



Universidad Autónoma de Madrid
School of Medicine

Doctoral Program of Molecular Biosciences

A Gene Regulatory Network to Control EMT Programs in Development and Disease

Hassan Fazilaty

Madrid, 2019

Department of Biochemistry
Faculty of Medicine
Universidad Autónoma de Madrid

Department of Developmental
Neurobiology
Instituto de Neurociencias de
Alicante CSIC-UMH

Doctoral Thesis

A Gene Regulatory Network to Control EMT Programs in Development and Disease

by

Hassan Fazilaty

Director

Professor M. Angela Nieto

Department of Developmental Neurobiology, Instituto de Neurociencias de Alicante
CSIC-UMH

Tutor

Professor Amparo Cano

Department of Cancer Biology, Instituto de Investigaciones Biomédicas 'Alberto Sols'
CSIC-UAM

April 2019, Madrid, Spain

Here, I certify that this work, presented by Hassan Fazilaty to obtain the PhD degree from Universidad Autónoma de Madrid, has been performed under my guidance and supervision.

Por la presente certifico que el trabajo que aquí se presenta para ser defendido por Hassan Fazilaty para la obtención del título de Doctor por la Universidad Autónoma de Madrid ha sido llevado a cabo bajo mi dirección. Para que así conste, lo firmo en Madrid a 25 de Febrero de 2019

M. Angela Nieto

To the love of my life, Soheila

and my beloved parents

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Prof. M. Angela Nieto for the continuous support of my Ph.D study and research, for her patience, motivation, enthusiasm, and immense knowledge. I have learned a lot from her, and her guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank my tutor Prof. Amparo Cano, who also supported me to do this Ph.D thesis in Universidad Autonoma de Madrid, as well as the rest of my thesis committee.

My sincere thanks also goes to Luciano Rago for his friendly mentorship and guidance since the first day when I started to learn the alphabets of science and experimental procedures, as well as Khalil K. Youssef with whom I experienced how to become a rigorous scientist.

I thank my fellow labmates in Instituto de Neuociencias de Alicante in the group of Cell movements in development and disease: Oscar Ocaña, Joan Galceran, Francisco Garcia-Asencio and Aida Arcas for their active collaborations, guidance and help during my research, as well as technical supports of Sonia Vega, Verona Villar, Hakan Coskun, Noemi Castroviejo, Diana Abad, Christina Lopez Blau and Sandra Moreno. I thanks the rest of lab members for their stimulating discussions, and for all the fun we have had in the last four years. Special thanks also goes to Auxi Casanova, who have always kindly assisted me with complicated administrative works starting even before I joined the lab.

First 3 years of my PhD research was supported by a Santiago Grisolia fellowship from Generalitat Valenciana (GRISOLIA/2014/004) (Project: GRISOLIA/2014/076 “Molecular Mechanisms of Tumor Progression”).

In addition I convey my gratitude to the scientific and executive members of the Instituto de Neurociencias, especially M^a Teresa García Hedo and former director Professor Juan Lerma, providing an opportunity to make this excellent environment for research and career development.

Last but not the least, I would like to thank my wife Soheila and my parents Hossein and Zohre to whom I owe everything, for supporting me spiritually throughout my life.

ABSTRACT

Abstract

Epithelial to Mesenchymal Transition (EMT) plays pivotal roles during development and diseases like cancer and fibrosis, through the activation of several EMT transcription factor (EMT-TF) families, including Snail, Zeb, Twist and Prrx. Prior data from the laboratory showed that the patterns of *PRRX1* and *SNAIL1* expression were complementary in chicken embryos and cancer cells, and that their functions in EMT subprograms, such as the regulation of stemness, also seemed to be distinct. Here, after examining zebrafish, chicken and mouse embryos, we find that this complementary expression of *Snail1* and *Prrx1* is conserved during vertebrate development. Moreover, analyzing public single-cell RNA sequencing databases of breast, head-and-neck cancer or melanoma patients and from mouse pulmonary fibrosis, we confirmed that this complementary expression is also present in pathological EMTs. By studying the transcriptome of cancer cells, gain and loss of function experiments for the two EMT-TFs, and the use of animal models, we describe a novel gene regulatory network (GRN) where *Snail1* and *Prrx1* form a double-negative feedback loop, involving miR-15 family. We have found that this GRN triggers an expression switch from *Snail1* to *Prrx1*, with *Snail1* being an early-response gene to EMT-inducing signals, which is followed by the activation of *Prrx1* that in turn attenuates *Snail1* expression through miR-15 family. We have also validated this GRN *in vitro* and *in vivo* highlighting its relevance in development and disease.

Resumen

La transición epitelio-mesenquimática (EMT) desempeña un papel fundamental durante la embriogénesis y la progresión de enfermedades como el cáncer y la fibrosis, mediante la activación de varias familias de factores de transcripción inductores de la EMT (EMT-TF), incluyendo Snail, Zeb, Twist y Prrx. Resultados previos obtenidos en el laboratorio mostraron que los patrones de expresión de PRRX1 y SNAIL1 eran complementarios en embriones de pollo y células cancerosas, y que sus funciones en subprogramas de la EMT, como la regulación de la pluripotencia, también parecían ser distintas. En este estudio, tras examinar el patrón de expresión de estos EMT-TFs en embriones de pez cebra, de pollo y de ratón hemos encontrado que la expresión complementaria de Snail1 y Prrx1 está conservada en los distintos vertebrados. Además, el análisis de bases de datos públicas obtenidos tras la secuenciación del transcriptoma de células individuales (scRNA seq) de pacientes con cáncer de mama, cuello y cabeza, melanoma, y de ratones con fibrosis pulmonar, confirmamos que esta complementariedad también existe en los procesos de EMT patológicos. Tras el estudio del transcriptoma de células cancerosas y experimentos de ganancia y pérdida de función para los dos EMT-TF, junto con el uso de modelos animales, hemos encontrado una red de regulación génica (GRN) donde Snail1 y Prrx1 forman un bucle de retroalimentación negativo, involucrando a la familia miR-15. Esta red promueve un cambio de expresión de Snail1 a Prrx1, siendo *Snail1* un gen de respuesta temprana a señales inductoras de EMT, que es seguida por la activación de Prrx1 que, a su vez, atenúa la expresión de Snail1 mediante la familia miR-15. Hemos validado este GRN *in vitro* e *in vivo*, revelando sus implicaciones en el desarrollo embrionario y la enfermedad.

INDEX

Table of Contents

ACKNOWLEDGEMENTS	v
ABSTRACT	vi
ABBREVIATIONS	xii
INTRODUCTION	1
1.1 Epithelial to mesenchymal transition (EMT)	3
1.2 EMT in embryonic development	7
1.3 EMT in cancer	10
1.4 EMT in fibrosis	13
1.5 Distinct roles of Snail1 and Prrx1 EMT-TFs	14
1.6 microRNAs and EMT	17
OBJECTIVES	21
MATERIALS AND METHODS	25
3.1 Cell culture	27
3.2 Microarray analyses	27
3.3 Plasmid and constructs	28
3.4 Treatments of cultured cells	28
3.4.1 Transfection of plasmids and interfering RNAs	28
3.4.2 Lentiviral infection	29
3.4.3 Dual luciferase reporter assays	29
3.4.4 TGF β administration	30
3.5 Total RNA extraction, cDNA synthesis and qPCR analysis	30
3.6 Chromatin immunoprecipitation (ChIP) assay	31
3.7 Mouse, chicken and zebrafish embryo sections	32
3.8 RNA and miRNA in-situ hybridization	33
3.8.1 Probe synthesis	33
3.8.2 Whole-mount mRNA visible and fluorescent <i>in situ</i> hybridization	33
3.8.3 Mature miRNA <i>in situ</i> hybridization	35
3.9 Western blot	36

3.10 Immunofluorescent (IF) staining.....	37
3.11 In silico analyses.....	38
3.11.1 Single cell RNA-seq public data	38
3.11.2 Meta-analysis of oncogenomic data	38
3.12 Statistical and data analysis	39
3.13 Tables.....	39
RESULTS	45
4.1 Prrx1 and Snail1 are expressed in complementary patterns in development and disease	47
4.1.1 Prrx1 and Snail1 complementary expression pattern is conserved in chicken and zebrafish embryos.....	47
4.1.2 Prrx1 and Snail1 complementary expression pattern is conserved in mouse embryos	49
4.1.3 Prrx1 and Snail1 expression pattern is mutually exclusive in cancer and fibrosis	51
4.2 Snail1 directly represses Prrx1 transcription, while Prrx1 induces its own expression.....	53
4.3 Prrx1 directly induces the transcription of miR-15 family members.....	57
4.4 miR-15 and Prrx1 share expression sites in embryonic development and correlate with survival in cancer patients	62
4.5 Prrx1 attenuates Snail1 expression through miR-15 family	64
4.6 Snail1 and Prrx1 are sequentially expressed during developmental EMT program	70
4.7 PRRX1 and miR-15 in the progression towards a full EMT.....	71
4.8 Snail1 but not Prrx1 is activated during partial EMT kidney fibrosis	75
DISCUSSION	77
5.1 The GRN that controls EMT programs	79
5.2 Differences and similarities in Snail1 and Prrx1-induced Programs.....	82
5.3 Prrx1 expression, a switch from partial to full EMT	84
CONCLUSIONS.....	87
REFERENCES.....	91

ABBREVIATIONS

3' UTR	3' untranslated region
ADAM	A disintegrin and metalloproteinase
AMF	Autocrine motility factor
BM	Basement membrane
BMP	Bone morphogenetic protein
CCND1	Cyclin D1
CCNE1	Cyclin E1
CDC25A	Dual specificity phosphatase
CDK	Cyclin dependent kinases
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
ChIP	Chromatin immunoprecipitation
CLL	Chronic lymphocytic leukemia
CTC	Circulating tumor cells
DEPC	Diethyl pyrocarbonate
DIG	Digoxigenin
DM	Dermomyotome
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial to mesenchymal transition
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FLOU	Fluorescent/Fluorescein
GRN	Gene regulatory network
H ₂ O	Water
H ₂ O ₂	Hydrogen peroxide
HCl	Hydrochloric acid
HIF	Hypoxia inducible factor
ILK	Integrin-linked kinase
KCl	Potassium chloride
KLF4	Kruppel-like factor 4

LPM	Lateral plate mesoderm
MET	Mesenchymal to epithelial transition
μg	Microgram
mg	Milligram
MgCl ₂	Magnesium chloride
miRNA	Micro-RNA
ml	Milliliter
mM	Millimolar
MMP	Matrix metalloproteinase
MNC	Migratory neural crest
MRE	Micro-RNA responsive element
mRNA	Messenger RNA
MTA3	Metastasis-associated protein 3
Mut	Mutant
NaAc	Sodium acetate
NaCl	Sodium chloride
nYFP	Nuclear yellow fluorescent protein
ORF	Open reading frame
OS	Overall survival
PAK1	P21-activated kinase
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
pH	Potential hydrogen
PNC	Premigratory neural crest
Pre-miRNAs	Precursor microRNAs
PSM	Pre-somitic mesoderm
PTH(rP)R	Parathyroid hormone related peptide receptor
RNA	Ribonucleic acid
RT-PCR	Reverse-transcription polymerase chain reaction
SC	Sclerotome
SCF	Stem cell factor
SDS	sodium dodecyl sulfate
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SSC	Saline-sodium citrate

tRNA	Transfer RNA
TBS	Tris-buffered saline
TEA	Triethanolamine
TFs	Transcription factors
TGF- β	Transforming growth factor- β
TIC	Tumor initiating capacities
TNC	Tenascin-C
WB	Wash buffer
WT	Wild type

Chapter 1

INTRODUCTION

1.1 Epithelial to mesenchymal transition (EMT)

The program that enables epithelial cells to acquire mesenchymal properties is known as *epithelial to mesenchymal transition* (EMT). It can be triggered in different contexts, from embryonic development to adult physiology and pathology. A polarized epithelial cell is normally sealed to a basement membrane (BM) and its adjacent cells. Cell-cell junctions comprise different types including adherens junctions, tight junctions, gap junctions, desmosomes, etc. and cell to BM adhesion takes place through integrin molecules along the basal surface. Therefore, epithelial cells are not able to migrate from their place of origin or invade and disseminate from their primary site. EMT is a conserved fundamental process that occurs primarily during embryonic development, enabling immotile epithelial cells move to different destinations to form a variety of organs in the developing embryo. The EMT is a crucial process during morphogenesis, as with the exception of the anterior central nervous system and the epidermis, all the tissues and organs result from cells that in origin underwent at least one EMT process. As such, the EMT is a dynamic and often transient process by which cells can undergo phenotypic changes towards the mesenchymal phenotype and then revert through a *mesenchymal to epithelial transition* (MET) in order to re-epithelialize and differentiate to fulfil their function. In adult physiology, EMT is reactivated during wound healing processes, where after injury, the epithelial tissue requires cells to migrate and seal the wounded area (Thiery et al., 2009, Nieto et al., 2016). In addition to normal developmental and physiological contexts, EMT is a major driving force in pathologies including cancer and fibrosis in which it can be aberrantly reactivated. Different states within the spectrum from the epithelial to the mesenchymal status can be acquired both during normal development and disease progression, usually leading to the delamination of cells from the tissue of origin, or invasion and metastatic dissemination in cancer, for which there is no efficient treatment presently (Nieto et al., 2016, Thiery et al., 2009). Although EMT has been studied extensively in the past decade, its complexity requires further studies so that we can get better insight into its fundamental physiological role and to tackle unmet clinical needs, the fight against cancer metastases and progressive fibrotic diseases.

The EMT is characterized by several cellular and molecular hallmarks, starting with EMT-inducing extracellular signals, the most potent of which are those triggered by the TGF- β superfamily. The cells initiate a complex signaling program that leads to

the activation of EMT-inducing transcription factors (EMT-TFs; e.g. Snail, Zeb, Twist and Prx factors) (Nieto and Cano, 2012b, Peinado et al., 2007) that lead to the loss of epithelial properties including adherens and tight junctions proteins (e.g. E-cadherin, claudins, occludins,...), and gain of mesenchymal properties including cytoskeletal proteins and the production of extracellular matrix components (ECM) (e.g. vimentin, fibronectin, collagen, ...) (Thiery et al., 2009, Acloque et al., 2009). These phenotypic changes allow otherwise immotile epithelial cells to disseminate ([Figure 1](#)).

With respect to E-cadherin, the most studied EMT-TF downstream target (Cano et al., 2000, Batlle et al., 2000), in addition to its transcriptional repression, *the* E-cadherin protein is subjected to internalization and degradation. This leads to the release of β -catenin from the membrane complexes, and its translocation to the nucleus, where it activates the transcription of additional genes that are involved in mesenchymalization of the cell (Behrens et al., 1996, Gilles et al., 2003). However, the transition from epithelial to mesenchymal states is not an ON/OFF switch, but rather consists of several intermediate steps that can be metastable or quasi-stable to be plastic and revert or to serve as final state in certain contexts, respectively. The intermediate states are referred to as partial-EMT states ([Figure 1](#)), which contribute to many of the EMT associated outcomes both during embryonic development and in adult physio-pathology.

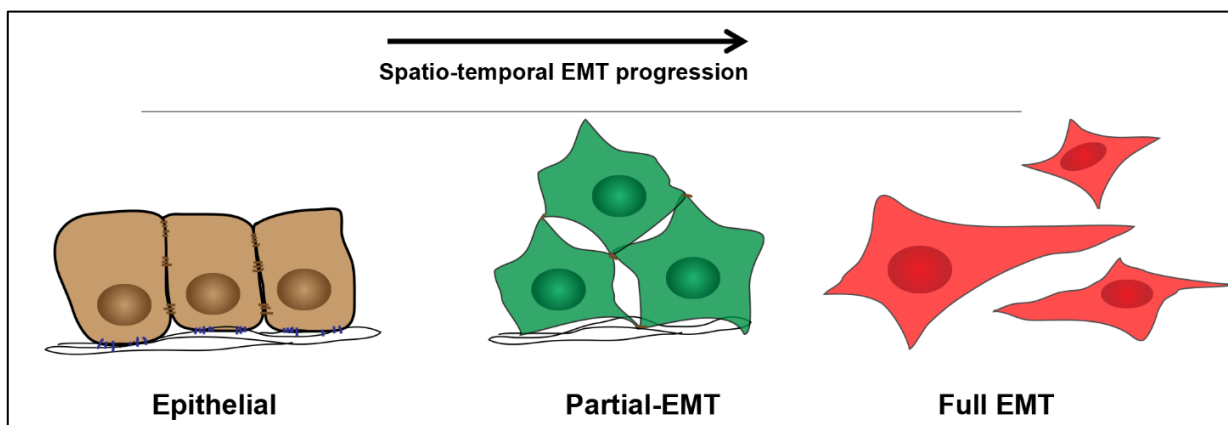


Figure 1. Schematic representation of cellular/morphological changes during EMT progression. Epithelial cells are normally adhered to each other and to the basement membrane through different cell-cell adhesion complexes and integrins, respectively. During EMT, when cells progress to a partial-EMT state, they start to lose cell-cell contact and their integrity within the tissue, while still preserve some cell-cell adhesion. When cells undergo a full EMT, they completely lose contact to each other and to the basement membrane and can be found as single mesenchymal cells.

The EMT-TFs, albeit all able to induce EMT, they act differently in terms of downstream targets and hierarchy of expression in different cellular contexts. These EMT-TFs are activated by different upstream signaling cascades, and are involved in many different gene regulatory networks (GRN) among themselves and also together with other regulatory molecules such as non-coding RNAs and protein-modifying enzymes (De Craene and Berx, 2013, Nieto and Cano, 2012a) Therefore, depending on the tissue, epithelial/mesenchymal cellular states, developmental stage and microenvironmental cues, different combinations of these EMT-TFs can be found, which leads to a complex set of EMT subprograms. Although all of these TFs induce transition of epithelial cells to mesenchymal derivatives, they are also involved in regulation of different associated programs (called subprograms or functions) which can be regulated in different manners and vary in potency. These subprograms include migration, invasion, stemness, survival, cell cycle attenuation, etc. (Ocana et al., 2012, Nieto, 2013). Thus, although all of these factors are known to induce EMT, one should be cautious about the context in which they are expressed in order to infer their detailed functional roles.

EMT-TFs are activated by different upstream signaling pathways, both in development and pathology. Those include members of the TGF β superfamily, Wnts, Notch, FGF, epidermal growth factor (EGF) and hypoxia inducible factor (HIF) among many others (Figure 2). For instance, Snail is rapidly induced upon TGF β and BMP administration *in vitro* and *in vivo* (Ocana et al., 2012), and is the first EMT-TF that is expressed during development in different processes including gastrulation and neural crest delamination (Nieto, 2002, Nieto et al., 1992, Nieto et al., 1994). Hypoxia inducible factor (HIF) is another upstream activator of EMT-TFs including Snail1, Snail2, Twist1 and Zeb1 that has been identified as one of the most critical EMT inducers as a result of intratumoural hypoxia, which leads to metastatic disease (Imai et al., 2003, Yang et al., 2008, Xu et al., 2015, Zhang et al., 2013, Zhang et al., 2015, Lu and Kang, 2010).

Adult organisms are still able to experience EMT for tissue homeostasis. In normal physiology, the different organs of the body consist of different types of epithelial cells, where EMT does not occur unless is needed to heal an injury or in pathology.

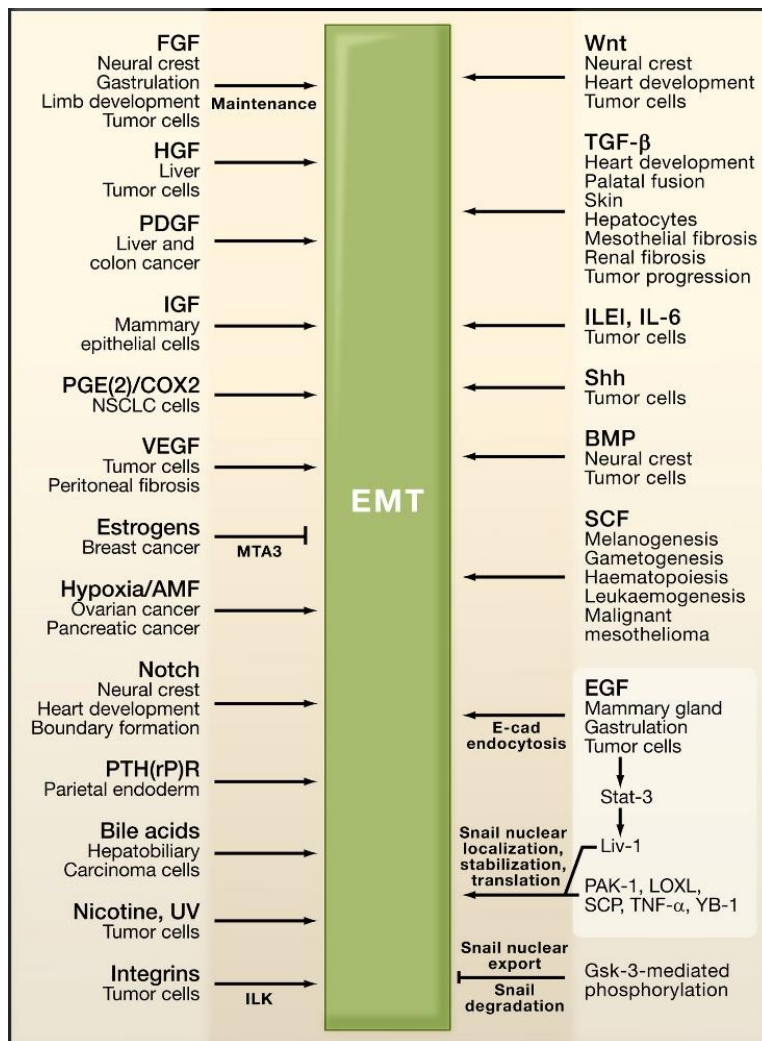


Figure 2. EMT induction plethora of signaling pathways. A variety of signaling pathways induce EMT in embryonic development and pathologies. Taken from (Thiery et al., 2009).

Upon formation of a lesion in an epithelial tissue, the inflamed microenvironment can activate EMT in neighboring epithelial cells to migrate and participate in the closure of the wounded area, a process called wound healing. For instance, in skin injuries,

keratinocytes at the border of the lesion activate the expression of the EMT-TF Snail2, which induces a migratory capacity in these cells to move not as single cells but rather as a collective migration resulting from an intermediate or partial EMT state induction (Arnoux et al., 2008, Savagner et al., 2005, Shaw and Martin, 2016). In this intermediate state, cells keep a loose connection among them as they do not completely shut down the expression of epithelial components, while they acquire migratory properties. When cells cover the wound, they undergo MET in order to re-epithelialize and form the normal tissue (Nieto et al., 2016). The wound healing process is an important homeostatic phenomenon, in which the EMT plays an important and beneficial role. The induction of this EMT is acute, but it can have deleterious effects when activated in a chronic manner during pathologies like cancer and fibrosis. For instance, a partial-EMT is the mechanism that leads to kidney fibrosis (Grande et al., 2015). Partial EMT states may help during tumor progression due to their intrinsic plasticity. As such, clusters of circulating tumor cells (CTCs) have increased metastatic potential (Aceto et al., 2014) and cancer cells with intermediate EMT phenotypes have

been described in mouse models and found in cancer patients (Taube et al., 2010, Sarrio et al., 2008, Pastushenko et al., 2018). Hence, a tight and dynamic regulation is needed for EMT to occur in a proper and timely manner depending on each context. Although there are differences between developmental, physiological and pathological EMT, the signaling networks and master regulators are common among them. Thus, to understand the biology of deregulated EMT in diseases, we can study it in development and physiology where they are regulated in robust circuits, then to take advantage of our knowledge as templates to comprehend pathological conditions.

1.2 EMT in embryonic development

The first EMT that occurs in an organism goes back to very early stages of embryonic development. Although differences exist among different species, the crucial role of EMT in morphogenesis is conserved throughout evolution (Thiery et al., 2009). EMT is induced during and even before implantation to form different extraembryonic cell populations. In mammals, among the very first EMT events is that occurring during the implantation process, in which the blastocyst adheres and invades the uterus resulting in homing and growth (Acloque et al., 2009). Evidence supports that extraembryonic cells express EMT markers and undergo morphological changes that help them to migrate and invade inside the uterus and pull the blastocyst (Knofler and Pollheimer, 2013, Pijnenborg et al., 1980, Pollheimer et al., 2014) ([Figure 3a](#)). Another EMT event occurs during the formation of parietal endodermal cells to form an extraembryonic cell population that will be incorporated into the yolk sac ([Figure 3b](#)). After taking part in formation of extraembryonic cells, EMT is induced in embryo proper to form different mesenchymal cells and their derivatives. Gastrulation and neural crest delamination are known as primary EMTs, as they are induced in epithelial cells that have never experienced an EMT before, and all the later events, including those EMT processes occurring during somites and heart development, are called secondary to tertiary EMT, as re-epithelialized cells undergo additional rounds of EMT to follow their cell fate and fulfill their corresponding function (Thiery et al., 2009). Thus, starting with gastrulation, the embryo will face several rounds of EMT in different tissues and at different developmental times. During gastrulation, ectodermal cells will migrate towards the primitive streak, where they will undergo EMT and delaminate generating the third embryonic layer, the mesoderm and the precursors of the definitive endoderm, both of

which will later give rise to a variety of tissues and organs (Thiery et al., 2009) (Figure 3c).

The induction and maintenance of the EMT during gastrulation occurs upon the activation of signaling pathways triggered by the transforming growth factor- β (TGF β) and fibroblast growth factor (FGF) families, respectively (Barrallo-Gimeno and Nieto, 2005, Ciruna and Rossant, 2001). The EMT-TFs that are activated during gastrulation and are conserved in different vertebrates are the members of the Snail family, the lack of which result in failure in gastrulation where mesodermal committed cells cannot downregulate E-cadherin and therefore are unable to migrate (Nieto et al., 1994, Carver et al., 2001). These mechanisms are evolutionary conserved, however, they are mediated by the use of different EMT-TFs in different species (Sefton et al., 1998), and occur in a short time-window, which require a tightly regulated activation/repression of gene transcription, as well as post-transcriptional regulation mechanisms.

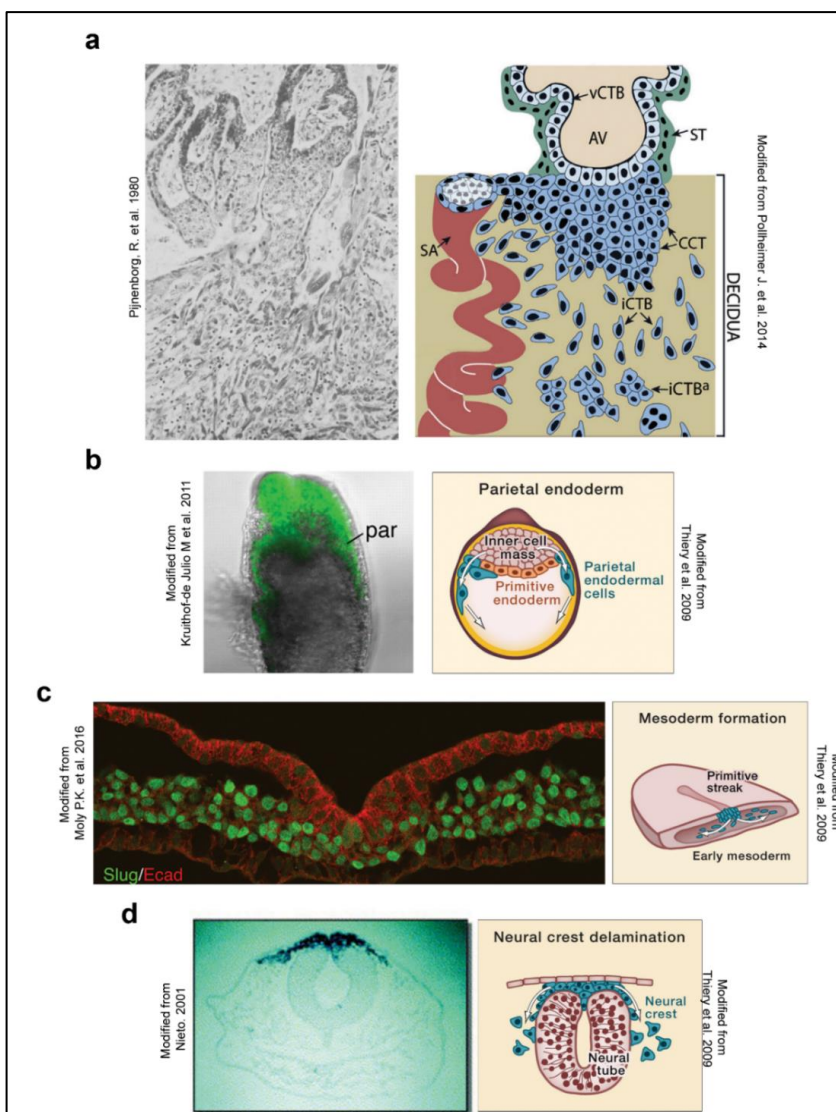


Figure 3. EMT during early embryonic development. (a) *Left*, Hematoxylin & Eosin staining of anchoring villus with trophoblast interstitial cells invading the decidua at 8 weeks of human embryonic development, taken from (Pijnenborg et al., 1980). *Right*, a scheme of the process. Extraembryonic cells that have undergone EMT invade the decidua (endometrium). Modified from (Pollheimer et al., 2014). (b) *Right*, scheme of the parietal endodermal cells derived from inner cell mass (that will give rise to the fetus), shown on

the *left* as green labeled cells migrating out. Modified from (Kruithof-de Julio et al., 2011) and (Thiery et al., 2009), respectively. (c) Slug (Snail2) expressing epithelial cells in the primitive streak of the gastrulating chicken embryo undergo EMT to form the mesendodermal cell layer. Modified from (Moly et al., 2016) and (Thiery et al., 2009), respectively. (d) *Snail2* expression in the delaminating neural crest cells in chicken embryo. Modified from (Nieto, 2001) and (Thiery et al., 2009).

A later EMT event transforms epithelial neural crest cells at the dorsal part of the neural tube, to delaminate and become migratory neural crest cells (Figure 3d). These migratory cells later differentiate to form different derivatives including most of the peripheral nervous system, the vast majority of pigment cells, cartilage and bone of the craniofacial structures, cardiac structures, smooth muscle cells, etc. These neural crest cells are exposed to a combination of signaling factors in a spatio-temporal manner, including bone morphogenetic protein (BMP), Wnt, etc. (Theveneau and Mayor, 2012).

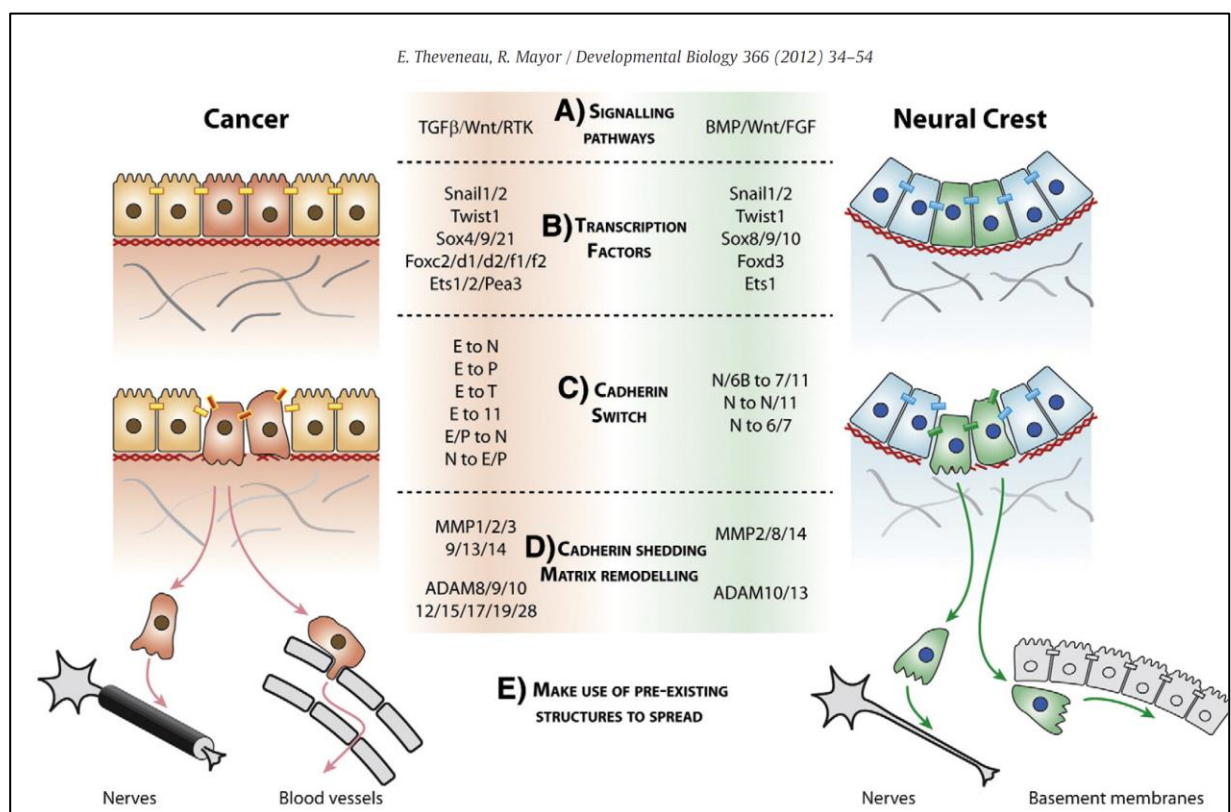


Figure 4. Similarities between neural crest delamination/migration and cancer metastasis. (A) Premigratory neural crest cells and early stage tumor cells show high levels of signaling activities that are similar, including TGFβ/BMP, Wnt and FGF/Receptor Tyrosine Kinase (RTK). (B) In both contexts, these signaling pathways induce the expression of EMT-TFs including Snail, Twist, etc. to initiate EMT. (C) Cell–cell adhesion loss allows neural crest

cancer cells to delaminate from their primary tissue. (D) Invasive neural crest and tumor cells secrete a variety of enzymes such as MMPs and ADAM families that in order to remodel the ECM and cellular adhesions to invade. (E) Neural crest and cancer cells migrate using pre-existing structures such as nerves and blood vessels in the case of tumor cells or nerves and the basement membrane of epithelia for neural crest cells. Taken from (Theveneau and Mayor, 2012).

During delamination, the majority of neural crest cells lose their cell-cell adhesion and apico-basal polarity, concomitant with forming blebs and lamelopodia (Ahlstrom and Erickson, 2009), all of which provide them with the ability to migrate. Several EMT-TFs are activated before, during, and after neural crest cell delamination, and feedback loops between them make the whole process to be robust (Calin et al., 2002, Martik and Bronner, 2017). From several perspectives, cellular and molecular mechanisms involved in embryonic neural crest delamination and migration have striking similarities to those occurring during cancer progression and metastasis, which include signaling pathways, EMT-TFs, switch in expression of proteins involved in cell junctional complexes and enzymes such as matrix metalloproteinases (MMPs) etc. (Figure 4) (Theveneau and Mayor, 2012). All these similarities make neural crest cells attractive templates to study cancer EMT.

1.3 EMT in cancer

Ectopic uncontrolled reactivation of EMT in adult tissues has been linked to pathological conditions including cancer progression. In these dysregulated states, constant expression of different combinations of EMT factors can lead to deterioration of the organ and/or misplaced migration and invasion as well as resistance to apoptosis, immune response and therapies. As during embryonic development, the EMT that occurs in each cancer type is different in terms of EMT-TFs expression depending on the tissue and state of mesenchymal induction, a notion called EMT-TF code (Nieto et al., 2016). These complexities make it difficult to the study and target EMT in these disease.

Cancer is one of the deadliest disorders in the world that makes it an essential target for research and development of therapies. More than 80% of cancers arise from epithelial tissues, which are called carcinomas, and amongst them, more than 90% of

deaths are associated to dissemination of the carcinoma (metastasis). Metastases are secondary tumors originated after the spread of tumor cells from the primary site to other organs, a condition for which no suitable treatments are currently available (Hanahan and Weinberg, 2011, Riggi et al., 2018). EMT provides the invasive and migratory properties to cancer cells in order to invade adjacent tissues and disseminate through circulation and to secondary organs (Nieto et al., 2016, Lambert et al., 2017), E-cadherin internalization or dislocation can also endow cells with the ability to migrate and invade, in this case in a collective way (Aiello et al., 2018). For metastatic colonization to occur, EMT should be reverted through the MET process in the distant organs (Ocana et al., 2012, Tsai et al., 2012). Therefore, targeting EMT can act like a double-edge sword, as it can impede cancer cells in the primary site from dissemination while it can be in favor already early disseminated cells (Harper et al., 2016, Hosseini et al., 2016) to revert and colonize at the secondary sites, thereby promoting metastatic colonization (Nieto et al., 2016). These information dictate us to deepen our knowledge before targeting EMT in cancer.

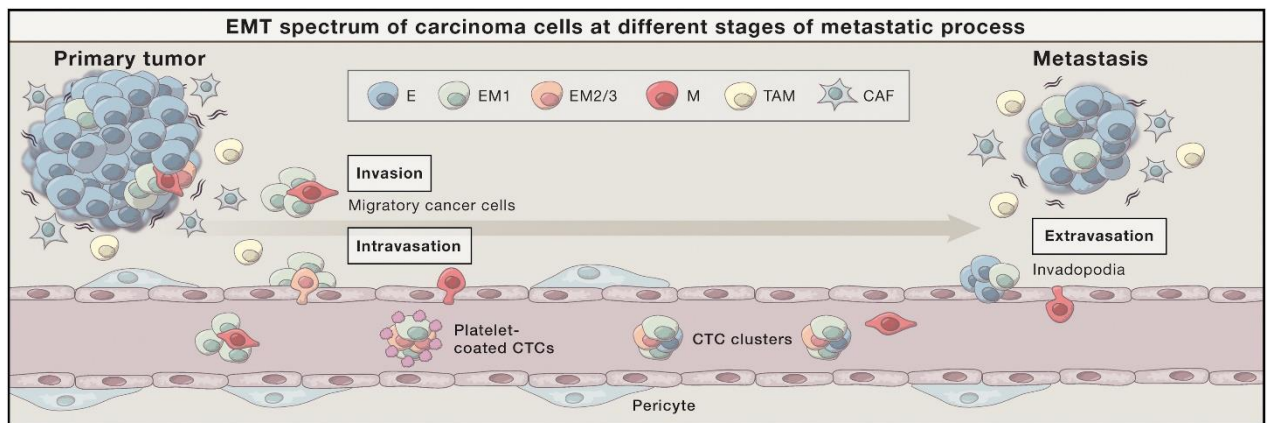


Figure 5. EMT during cancer metastasis. Tumor cells with a gradient epithelial to mesenchymal states can be found in different regions of the tumor. EMT induces invasive and migratory abilities in tumor cells, which enable them to disseminate from the primary site, intravasate into the circulation as CTCs in their way to colonize the distant organs after extravasation from the circulation. Taken from (Nieto et al., 2016).

Carcinomas are comprised of highly heterogeneous cell populations with different states of epithelial and mesenchymal properties among cells residing in different regions of the tumor. In the invasive front of the tumors, mesenchymal cells can be found in close contact with other microenvironmental cells and connective

tissue, called stroma, while in more central regions cells with intermediate phenotype might be found among fully epithelial cells (Nieto et al., 2016) (Figure 5). Analyses of circulating tumor cells (CTCs) and primary tumors in human patients provide evidence of EMT induction during cancer progression (Yu et al., 2013, Salnikov et al., 2012, Ocana et al., 2012, Ye et al., 2015). Accordingly, several mouse models have shown that EMT occurrence plays an essential role in the first step of the metastatic cascade, as after deleting the specific context-dependent EMT-TF, such as Snail, Zeb or Twist, tumors show significantly reduced metastatic potentials (Ye et al., 2015, Krebs et al., 2017, Xu et al., 2017, Tran et al., 2014). In several models, different degrees of epithelial plasticity have been linked to metastatic potential in primary tumors (Pastushenko et al., 2018, Latil et al., 2017), and single cell RNA-sequencing (scRNA-seq) of human cancer patients has identified the existence of these partial-EMT cells that contribute to metastatic colonization (Puram et al., 2017) (Figure 6).

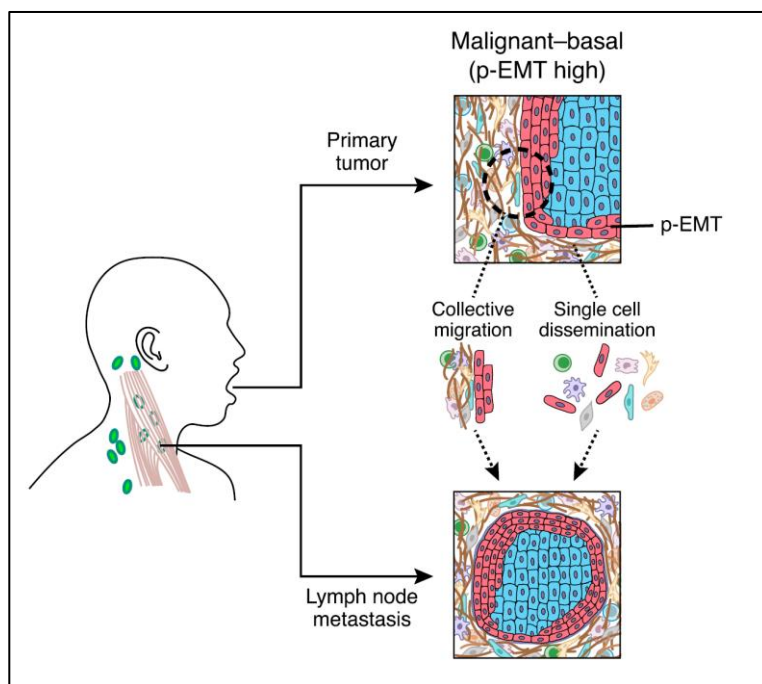


Figure 6. Partial EMT (p-EMT) cell population in cancer patients contribute in metastatic colonization.

Tumor cells in close contact with stromal cells show p-EMT properties and take part in invasion and metastasis of human malignant-basal head-and-neck squamous cell carcinoma tumors. Taken from (Puram et al., 2017).

In addition to invasive and migratory capabilities, EMT grants cancer cells with resistance to therapy and cell death. EMT inducers such as Snail increase survival of cells by affecting several downstream targets including p53, caspase proteins, ERK and PI3K signaling pathways (Peinado et al., 2007, Barrallo-Gimeno and Nieto, 2005). Inhibiting cell death and inducing survival pathways is concomitant with blocking cell cycle (Vega et al., 2004) through which EMT inducers also confer chemoresistance during cancer progression (Fischer et al., 2015, Zheng et al., 2015), likely because

available therapeutic agents can effectively target cycling but not mesenchymal non-cycling tumor cells. Thus, EMT needs to be considered as an essential mechanism during cancer metastasis as it not only makes tumor cells able to invade and migrate from their primary tissue but also confers survival properties to resist cell death and therapy.

1.4 EMT in fibrosis

Fibrosis is a type of degenerative disorder in which an organ gradually loses its function and eventually fails, which accounts for an important public health issue (Hutchinson et al., 2015). Fibrosis refers to an accumulation of myofibroblasts in an inflamed microenvironment in tissues, which results in secretion of an excessive amount of collagen deposited as fibers leading to tissue thickening. In the fibrotic disease, resident and recruited fibroblasts are differentiated into myofibroblasts that can generate collagen network. Fibrosis occurs in many different organs including kidneys, liver, lung, heart etc. EMT is believed to be a key process to initiate these aberrations (Thiery et al., 2009). TGF β signaling is one of the major inducers of EMT in fibrosis as well (Yang et al., 2010), which for instance in kidney fibrosis results in reactivation of Snail1 and Twist1 EMT-TFs in renal epithelial cells that lead to organ failure (Boutet et al., 2006, Kida et al., 2007). Notably, the EMT that occurs during renal fibrosis is a partial-EMT in which renal tubules get deteriorated concomitant with parenchymal damage, but cells do not disseminate from their primary sites but stay in contact. Importantly, these damaged cells secrete cytokines and chemokines to promote both fibrogenesis and inflammation. As such, in the absence of EMT reactivation in renal epithelial cells, fibrosis cannot progress (Grande et al., 2015, Lovisa et al., 2015). This works in favor of treatment, as inhibiting Snail1 when fibrosis is already established has been shown to impede the progression and even reverse the disease (Grande et al., 2015) ([Figure 7](#)). Therefore, unlike cancer that repression of EMT can be in favor of metastatic colonization, targeting EMT in fibrosis is beneficial.

These complex transformation of the cell during the EMT process, both in development and disease, is performed by master transcription factors through different combinations. Thus, understanding the interaction between these transcription factors may be the key to fully understand physiological EMTs and to control EMT in cancer and fibrosis.

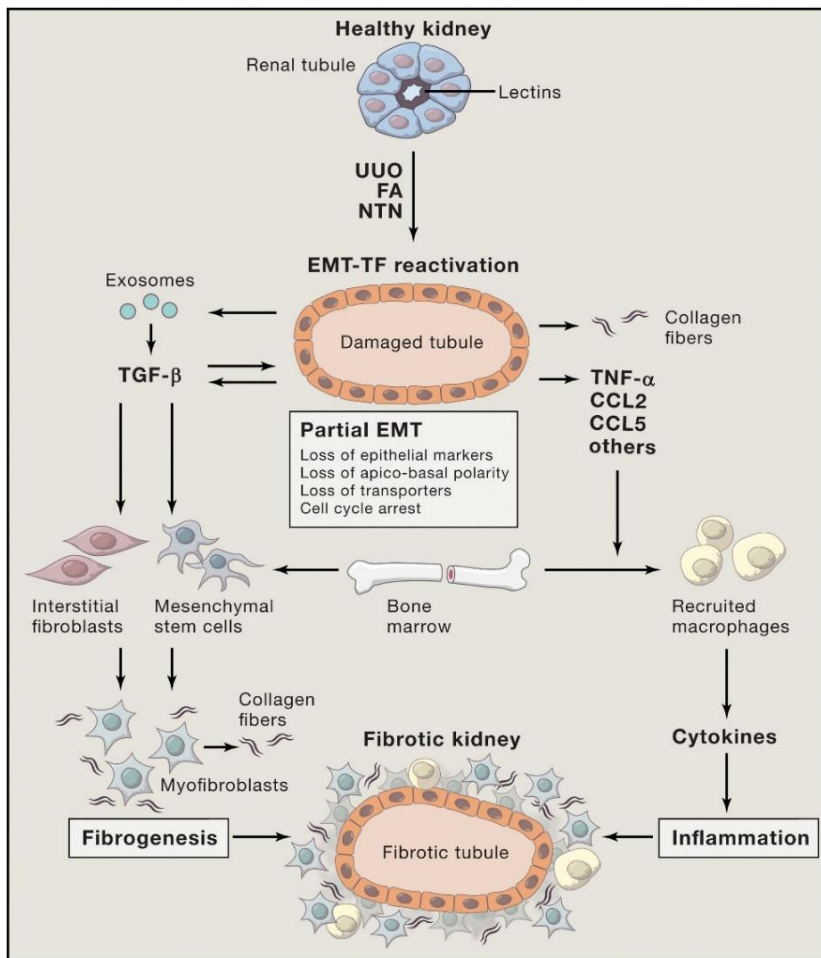


Figure 7. Partial EMT in renal fibrosis. Upon induction of fibrotic disease in the kidney, EMT-TFs including Snail1 and Twist1 can be reactivated, lead to damaged renal tubules that lead to renal insufficiency. In addition, they secrete several cytokines and chemokines that lead to a fibrotic and inflamed stroma and eventually organ failure. Taken from (Nieto et al., 2016).

1.5 Distinct roles of Snail1 and Prrx1 EMT-TFs

Different EMT-TFs trigger different subprograms. In cancer, Snail1 expression is associated with malignant phenotype and poor prognosis (Barrallo-Gimeno and Nieto, 2005, Ye et al., 2015), while the expression of Prrx1 is associated with good prognosis and metastasis-free disease because even though it promotes delamination from the primary tumor and invasion, cancer cells need to downregulate it to colonize and develop metastatic outgrowths (Ocana et al., 2012). Snail1 and Twist1 induce stemness (Mani et al., 2008), while Prrx1 expression is concomitant with the loss of stemness (Ocana et al., 2012). This also explains why it needs to be downregulated at the metastatic site, in order for cancer cells to recover tumor initiating capacities (TIC). These differences are already evident during embryonic development, where *PRRX1* and *SNAIL1* are expressed in a complementary manner in several regions, including the neural crest and the lateral plate mesoderm in the chicken embryo (Ocana et al.,

2012). However, as every mesodermal cell already underwent Snail-dependent EMT during gastrulation and neural crest delamination, and at later developmental stages other EMT-TFs are expressed, this suggests a tightly regulated crosstalk between Snail1 and other EMT-TFs including Prrx1, not only in development but also in pathological EMTs, that might impose EMT-TF specific programs.

Snail genes encode evolutionary conserved transcription factors (Manzanares et al., 2001) that play central roles in all contexts of EMT and were the first to be identified (Nieto et al., 1994, Nieto et al., 1992). Snail1 plays many roles in regulating cellular states by targeting numerous downstream effectors and other TFs, as result of itself being activated by a variety of signaling networks ([Figure 8](#)) (Barrallo-Gimeno and Nieto, 2005, Wu and Zhou, 2010). Notably, Snail is a potent repressor of epithelial components (Boutet et al., 2006, Cano et al., 2000, Grande et al., 2015). Snail also attenuates the cell cycle by repressing essential components including cyclin dependent kinase (CDK) and cyclins (Vega et al., 2004), with the cells concentrating their energy for invasion and migration rather than proliferation. Although generally known as an inhibitor, Snail can also behave as an activator when interacting with Twist1 (Rembold et al., 2014) or CBP (Hsu et al., 2014). It also activates Zeb1 through which induces tumor initiating capacities during cancer progression and metastasis (Ye et al., 2015). As a potent survival effector, Snail represses apoptosis triggered by TGF β , playing an essential role in switching TGF β signaling from tumor suppressor to tumor progression (Franco et al., 2010, Caja et al., 2011, Fabregat et al., 2016). In development, Snail plays essential roles in many other processes other than its critical roles in gastrulation and neural crest delamination. For instance, Snail is a regulator of left-right asymmetry and heart positioning during development of chicken and mouse embryos (Patel et al., 1999, Murray and Gridley, 2006, Ocana et al., 2017). Thus, Snail is not only an essential EMT inducer, but also a pivotal player in many derived processes, therefore understanding its regulation is essential for understanding the EMT-related programs.

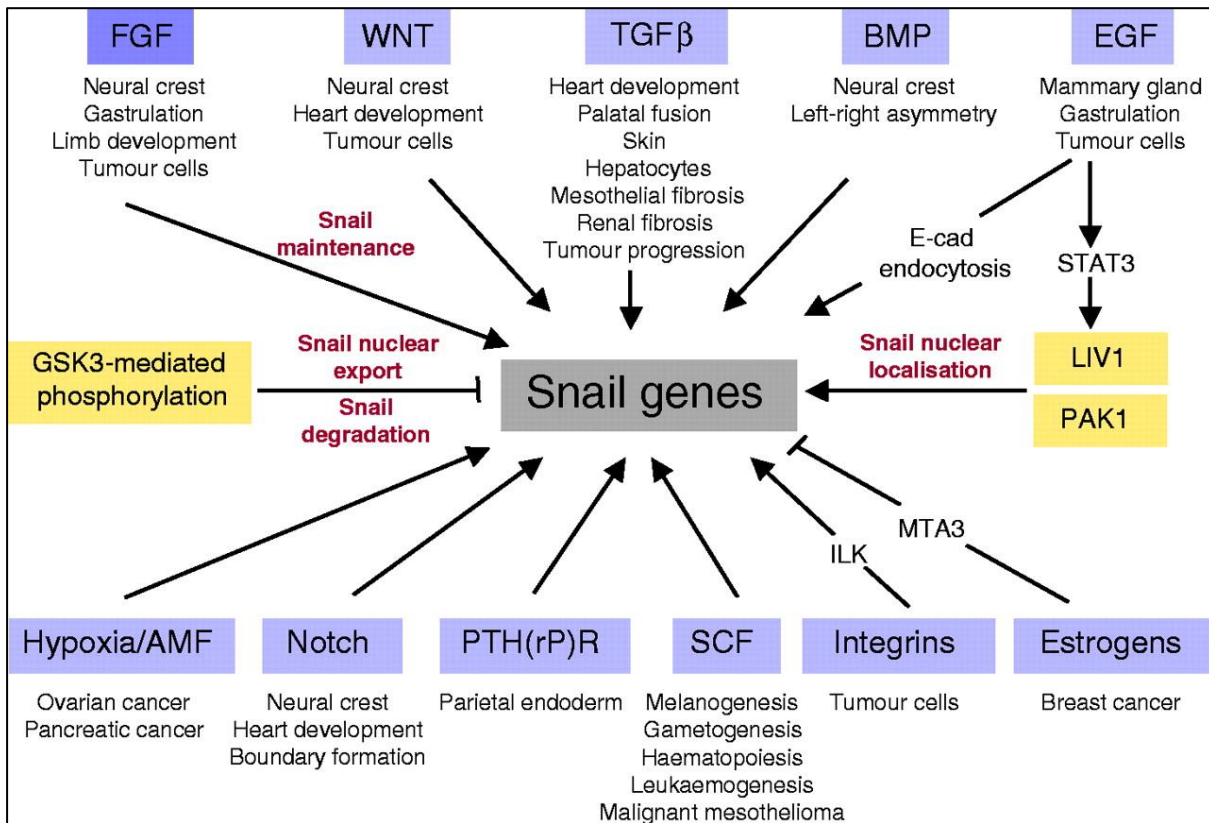


Figure 8. Snail is a central transcription factor induced during EMT. Snail factors are a target of many signaling pathways and regulate numerous cellular processes during embryonic development and different types of cancer. Besides being regulated by many signaling cascades at the transcriptional level, Snail1 protein activity is also tightly controlled by posttranscriptional modifications and its subcellular localization, for instance through kinases like GSK3 and PAK1, and by the zinc-finger transporter LIV1. AMF, autocrine motility factor; ILK, integrin-linked kinase; MTA3, metastasis-associated protein 3; PAK1, p21-activated kinase; PTH(rP)R, parathyroid hormone related peptide receptor; SCF, stem cell factor. Taken from (Barrallo-Gimeno and Nieto, 2005).

The paired-related homeobox protein, Prrx1 is also an EMT-TF with important roles both in embryonic development and in disease (Ocana et al., 2012). Prrx1 is expressed in the mouse embryo mesenchyme of facial, limb, and vertebral skeletal progenitors, and its loss during development is lethal soon after birth due to deficiencies in morphogenesis of those structures (Martin et al., 1995). Heart-positioning in vertebrates is controlled by a left-right asymmetric EMT process that generates asymmetric forces more potent on the right side that displace the posterior pole of the heart to the left (Ocana et al., 2017). Interestingly, while the cellular process is conserved, the EMT-TF involved is different in different species. In the chicken

embryos, both Snail 1 and Prrx1 contribute to heart laterality from different territories, dorsal and ventral, respectively. In the fish Prrx1 is expressed in both territories and the same applies to Snail in the mouse (Ocana et al., 2017). Additionally, Prrx1 is required for lung vascularization mediated by tenascin-C (TNC), an ECM component that regulates endothelial cells differentiation and therefore, functional vascular networks (Ihida-Stansbury et al., 2004). As a direct activator of TNC, Prrx1 regulates the migration of fibroblasts by formation of TNC-enriched ECM (McKean et al., 2003). In cancer, Prrx1 cooperates with Twist1 in inducing mesenchymal phenotype, both of which need to be downregulated in the metastatic site for a successful metastatic colonization (Ocana et al., 2012). Altogether, Prrx1 is an EMT-TF both in development and disease and its function being distinct from Snail1 may be one of the keys for deeper understanding of the EMT process (Figure 9).

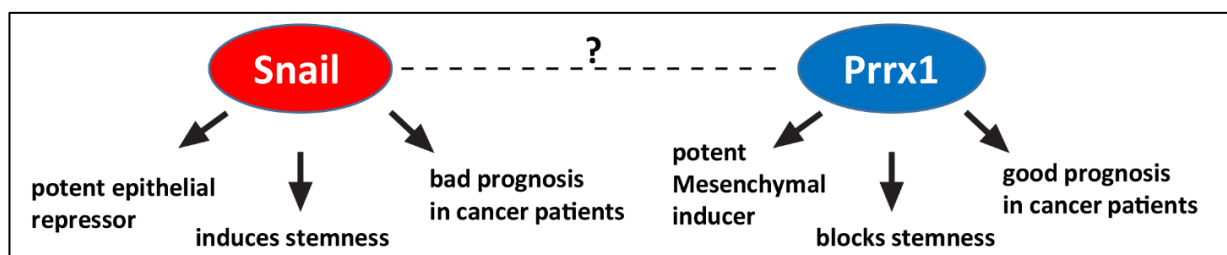


Figure 9. Is there any regulatory mechanism between Snail1 and Prrx1 EMT-TFs? Considering that both Snail and Prrx1 are EMT-TFs with different functioning in subprograms of EMT both in development and disease, it arises the question of whether they regulate each other as they normally cannot be dominant in a given context both together.

1.6 microRNAs and EMT

Micro-RNAs (miRNAs) are short non-coding regulatory RNAs that post-transcriptionally attenuate gene expression. miRNAs control the essential cellular functions during embryonic development, homeostasis and many different pathologies (Rodriguez-Aznar et al., 2013, Pernaute et al., 2011, Shimono et al., 2009). They function by binding to their complementary sequences in target mRNAs. When bound, miRNAs block the translational machinery or lead to mRNA degradation. miRNAs are between 18-22 nucleotides long and are transcribed from a variety of regions in the genome, including coding sequences and introns of genes or non-coding regions. After

transcription, miRNAs undergo several rounds of processing from primary-miRNA (Pri-miRNA) to precursor-miRNA (Pre-miRNA) and eventually mature miRNA (Figure 10).

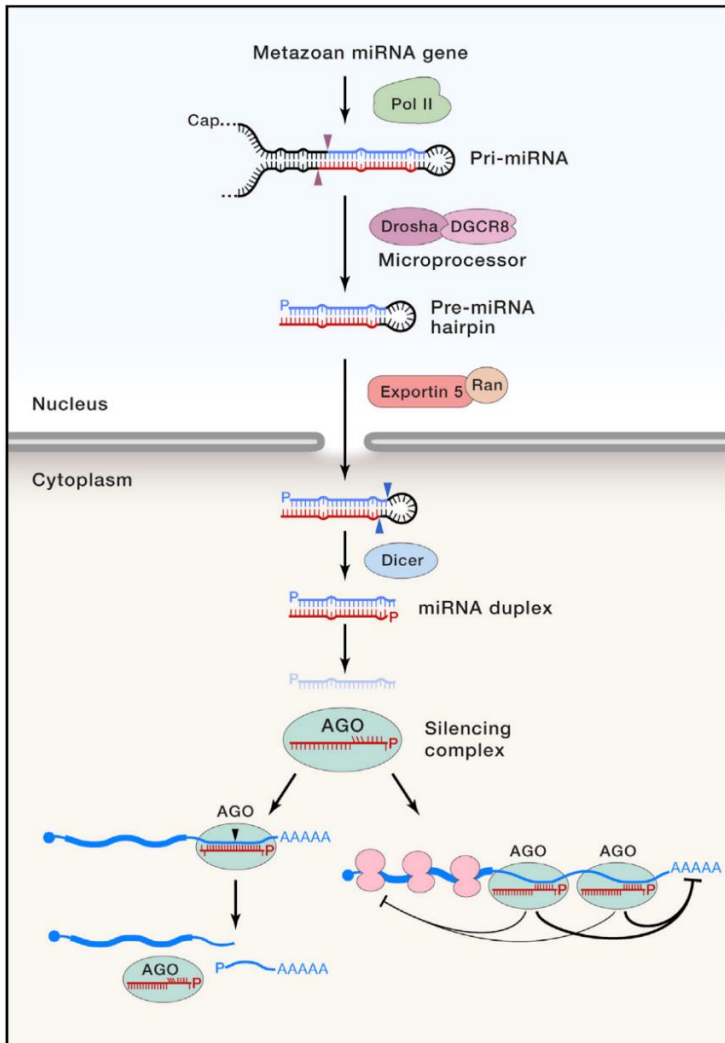


Figure 10. microRNAs biogenesis and function. After transcription in the nucleus and the processing of Pri-miRNAs to Pre-miRNAs, they get transported to the cytoplasm and after maturation they make a complex with several proteins called RISC or silencing complex that can target mRNAs for degradation, if they perfectly match with the target sequence, or silence translation when there is mismatch in the binding. Modified from (Bartel, 2018).

Several protein complexes control the processing machinery. Starting inside the nucleus Pri-miRNAs are trimmed and form Pre-miRNA that are exported to the cytoplasm and eventually get incorporated as single stranded into a complex together with several proteins called RNA Induced Silencing Complex or RISC. This complex scans different mRNAs to find partially complementary sequences (usually in the 3' untranslated region, 3'UTR) that matches the sequence in the mature miRNA. This complementarity can be perfect (like in plants) or imperfect (Figure 10). The most essential region in miRNAs is the “seed sequence”, usually from nucleotides 2 to 7 of miRNA that best matches the target sequence. The seed sequence is also the most conserved region in miRNAs, and different miRNA molecules with the same seed sequence are known as family members, as they can potentially target the same mRNAs. Several miRNA molecules are able to target the same mRNA, and a single

miRNA can target many different transcripts, a pleiotropic function that makes them powerful regulators of many cellular processes including EMT (Bartel, 2018, Bartel, 2009).

Several miRNAs regulate EMT at different levels, including targeting EMT-TFs, as well as being targeted by EMT-TFs themselves. Well studied examples include Snail1 and Zeb1 that are attenuated by miRNAs of the miR-34 and miR-200 families, respectively, while they also repress these miRNAs and form a reciprocal negative feedback loop (Figure 11). This loop plays a pivotal role in regulating the balance between the protection of the epithelial state and the EMT (Cano and Nieto, 2008, Gregory et al., 2008, Kim et al., 2011b, Bracken et al., 2008, Burk et al., 2008). Other miRNAs known to regulate EMT include miR-30, miR-29, miR-10, miR-1, miR-183-182 etc. that by targeting different components can either promote or repress the EMT process (Diaz-Lopez et al., 2014) (Figure 11). Twist1 induces miR424, resulting in the induction of a mesenchymal program and partial EMT in breast cancer metastasis (Drasin et al., 2015a). Thus, the regulation of many developmental and pathological EMT are tightly linked to different miRNAs, by which an accurate and time-dependent fine-tuning of target genes can be programmed. However, miRNAs that can act as *bona fide* regulators of Prrx1 are not known.

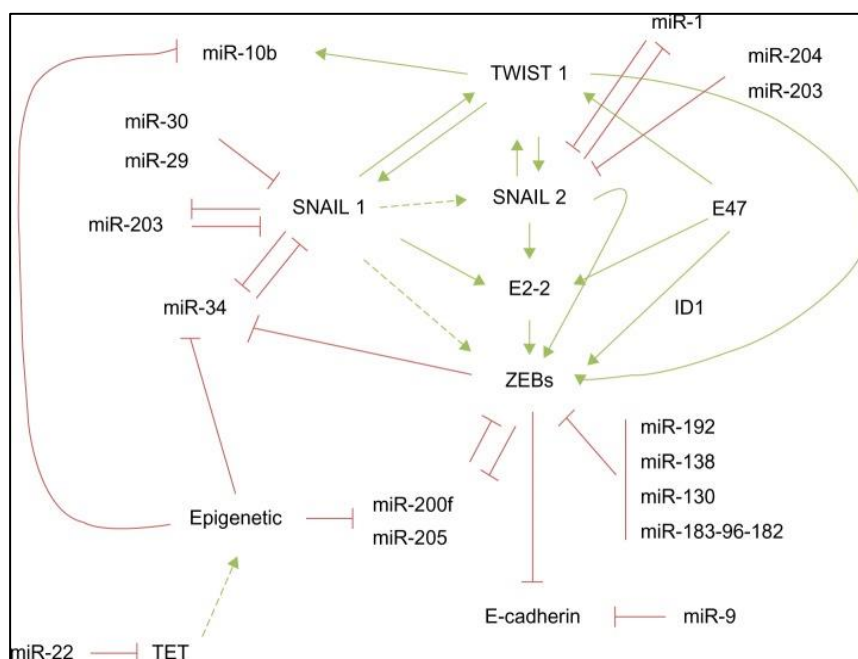


Figure 11. MicroRNAs regulate EMT-TFs. Many microRNA families including miR-200 and miR-34 regulate the expression of EMT-TFs like Zeb and Snail families, while reciprocally they get repressed by these EMT-TFs forming double-negative feedback loops. Taken from (Diaz-Lopez et al., 2014).

Chapter 2

OBJECTIVES

2. OBJECTIVES

Snail1 and Prrx1 are the most dissimilar EMT-TFs when considering both embryonic development and cancer progression. They exhibit complementary expression patterns during development of chicken embryos, although both are induced by the same signaling factors i.e. TGF β family. The EMT induced by Snail1 and Prrx1 is also different in terms of the cellular behavior they impinge *in vitro* (unpublished data from the lab) and also in tumor initiating capacities. As such, even though the two of them induce EMT, motility and invasion, the expression of Snail1 and Prrx1 has been respectively associated with bad and good prognosis after the analysis of databases from cancer patients.

Hypothesis

Understanding the mechanisms that regulate Snail and Prrx in different contexts is crucial to understand the complexity of EMT in health and disease

Objective

The objective of this thesis is to identify the existence, relevance and importance of molecular regulatory mechanisms that direct cells to undergo different EMT programs. As our aim is to find commonalities and to reach robust conclusions on the mechanisms that regulate EMT, we set out to analyze different vertebrate embryos and different pathological states. Specifically we have:

- Studied the expression patterns of Snail1 and Prrx1 in chicken, zebrafish and mouse embryos, as well as in cancer and fibrosis.
- Investigated the relationship between Snail1 and Prrx1 in terms of transcriptional regulation.
- Validated the existence and relevance of such molecular mechanisms *in vitro* and *in vivo*.

Chapter 3

MATERIALS AND METHODS

3.1 Cell culture

MDA-MB-231 (MDA231), MDA436, BT549 and SUM149PT human breast tumor cell lines, HEK293 human embryonic kidney and MDCK-II and MDCK-NBL dog kidney epithelial cells, and NIH3T3 mouse fibroblast cells were purchased from the ATCC (Virginia, USA). MDA231, MDA436 and BT549 human tumor cell lines were cultured in DMEM:F12 HAM media (1:1), and SUM149PT cells in F12 HAM media, supplemented with 10% heat inactivated fetal bovine serum (FBS) (Sigma), 10 µg/ml insulin (Roche), 1% Gentamicin (Sigma) and 1% amphotericin (Sigma). HEK293 and MDCK cells were cultured in DMEM supplemented with 10% heat inactivated FBS (Sigma) and 1% Gentamicin (Sigma) 1% amphotericin (Sigma). NIH3T3 cells were cultured in DMEM supplemented with 10% heat inactivated calf new-born serum (NBS) (Sigma), 1% amphotericin (Sigma) and 1% Gentamicin (Sigma). Cells were kept at 37°C and 5% CO₂, and the media was replaced every two/three days. HEK293 and MDCK cells were passaged when they reached 80% confluency 1:10 every 48 hours, while BT549 and MDA231, MDA436 and SUM149PT cells were passaged when reached 80% confluency 1:5 every 72 hours. Cells were discarded after 5 consecutive passages and replaced by freshly thawed stocks.

3.2 Microarray analyses

MDA231 human tumor cell line was transduced using lentiviral system in order to stably express PRRX1 (MDA231-P) (Ocana et al., 2012). The microarray analyses were done by Aida Arcas. Three independent samples per condition, MDA231 control and MDA231-PRRX1 cells, were hybridized to Human Gene 2.0 ST expression arrays (Affymetrix). Microarray data was analyzed using R and Bioconductor (Gentleman et al., 2004). Expression data was preprocessed and normalized using the RMA algorithm (Irizarry et al., 2003), and differential expression analysis was performed using limma (Smyth, 2005) t-test to extract statistically significant changes between PRRX1 and control samples. Accordingly, the genes that have a fold change > 1.5 and corrected P-value < 0.05 (Benjamini–Hochberg FDR) were considered as being differentially expressed.

3.3 Plasmid and constructs

For overexpression studies, human PRRX1 ORF ([ENSG00000116132](#)) was cloned in the pBABE expression vector, and mouse *Prrx1* ORF ([ENSMUSG00000026586](#)) was cloned in the pCDNA3.1 expression vector. Lentiviral vectors were generated via subcloning nuclear yellow fluorescent protein (nYFP)-P2A-*Prrx1*-L into pTRIPZ construct. P2A is a self-cleaved short peptide derived from porcine teschovirus-1 which is cleaved upon translation resulting to an intact protein sequence plus reporter in an equivalent manner (Kim et al., 2011a). For RNA interference, nYFP was subcloned into pLKO.1 construct containing PRRX1 specific shRNA (Dharmacon, RHS3979-201751764, Mature antisense: TTATTGGCTAGCATGGCTCTC) (Ocana et al., 2012). For miRNA overexpression studies, approximately 150 base pairs up- and downstream of the stem-loop sequences of each miRNA pair (obtained from www.mirbase.org) were cloned into the pCDNA3.1 vector. 3' UTRs of mouse, human, chicken and zebrafish *Snail1* were cloned downstream of Firefly luciferase ORF in pGL3 vector. Promoters of human and mouse *Mir-15-P1/2d* as well as human and mouse *Prrx1* were cloned upstream of the Firefly luciferase ORF in pGL3 vector. For mutagenesis, primers were designed using the online tool <http://bibiserv.techfak.uni-bielefeld.de/>. PCR was performed using Pwo polymerase following manufacturer's instructions, and PCR products were treated with DpnI restriction enzyme, then transformed in bacteria to be amplified, extracted using Mini-prep and then sequenced to confirm the mutation/deletions. All constructs were generated using the primers described in [Table I](#).

3.4 Treatments of cultured cells

3.4.1 Transfection of plasmids and interfering RNAs

BT549 cells were transfected with either control pLKO.1 vector or PRRX1 specific shRNA (Ocana et al., 2012) using Lipofectamin 3000 (Invitrogen) following the manufacturer's instructions. 250,000 cells were seeded in 6-well plates 24 hours prior transfection, and 48 hours after transfection cells were lysed for RNA extraction. A nYFP was subcloned into the pLKO.1-PRRX1 shRNA vector for the purpose of enriching the transfected cells by sorting the cells using Fluorescence-activated cell sorting (FACS) (BD FACSARIA III, USA). Sorted cells were directly collected into lysis

solution from *mirVana*TM miRNA Isolation Kit and subjected to RNA extraction (see below).

MDCK-NBL cells were transfected once with siRNA against dog PRRX1 24h prior to TGF β treatment, and a second round of transfection was done 4 days after the treatment (see below), using Lipofectamin RNAi MAX (Invitrogen). Block-itTM (Invitrogen) fluorescent oligo reagent was used as control. 3 different siRNA against the dog PRRX1 sequence were designed and tested, and the most efficient one in terms of reducing the levels of *PRRX1* transcripts was chosen for subsequent experiments (cfaPRRX1 siRNA: Sense: GAGCGCGUCUUUGAGAGAACACACU, antisense: AGUGUGUUCUCUCAAGACGCGCUC). 10,000 cells were seeded 24h prior transfection and then were collected 8 days after TGF β treatment for RNA extraction.

3.4.2 Lentiviral infection

MDA231 cells were infected by pTRIPZ inducible lentiviral system containing nuclear nYFP-P2A-Prrx1 or control vector with turbo RFP. The pool of infected cells was selected by treatment with puromycin for 48h. Infected cells were seeded in 10 cm dishes, subjected to 2 μ g/ml doxycycline to induce the Tet-ON promoter (tetracycline-controlled transcriptional activation that is induced upon presence of antibiotic tetracycline or one of its derivatives, doxycycline) and consequently FACS-sorted after 48h for either YFP or RFP positive signals for Prrx1 overexpressing or control cells, respectively. Cells were collected immediately after sorting in lysis solution from *mirVana*TM miRNA Isolation Kit (Invitrogen).

3.4.3 Dual luciferase reporter assays

HEK-293 and BT549 cells were cultured in 24-well plate 10h before transfection. 100,000 cells/well were seeded in triplicate and transfected with 100 ng of the constructs containing sequences for promoters or 3'UTRs cloned in pGL3P vector, 50 ng of the pRL control vector (Renilla luciferase) and 200 ng of the pcDNA3.1 plasmid subcloned with different transgenes for Pre-miRNA or Prrx1/Snail1 overexpression, using the Lipofectamin 2000 (Invitrogen). The media was refreshed 12h after transfection. After 48h, the cells were lysed using 5x lysis buffer (Promega), and the

enzymatic activity was analyzed after adding Luciferase Assay and Stop&Glo reagents (Promega), to measure Firefly and Renilla successively, all using Sirius Luminometer (Berthold, Germany).

3.4.4 TGF β administration

10,000 MDCK-II or MDCK-NBL cells were seeded on 6-well plates 24h before administration of 5 ng/ml TGF β (TGF β 1, Human, Millipore, USA). Cells were collected at different time-points, 1 hour (1h), 4h, 1 day (1d), 2d, 4d, 8d and lysed for RNA extraction. For immunofluorescence analyses, cells were seeded on top of coverslips. TGF β was refreshed every 48h.

3.5 Total RNA extraction, cDNA synthesis and qPCR analysis

For gene expression assays, total RNA was extracted using the illustra RNeasy Mini isolation kit (GE Healthcare), following the manufacturers' instructions. For Reverse transcription cDNA synthesis, oligo (dT)₁₈ and random hexamer primers with the Maxima First Strand cDNA Synthesis kit (Thermo Scientific) were used, following the manufacturers' instructions. Primers used are listed in [Table I](#). Quantitative RT-PCR was performed using Fast SYBR Green Mastermix (Applied Biosystems), in a Step One Plus machine (Applied Biosystems) according to the manufacturers' instructions. Relative levels of expression were calculated using the comparative Ct method normalized to the internal control *TBP* housekeeping gene.

For mature miRNA expression assays, total RNA enriched for small RNAs was extracted using *mirVana*TM miRNA Isolation Kit (Invitrogen), following the manufacturers' instructions. Reverse transcription reactions were performed using TaqMan[®] MicroRNA Assay (Applied Biosystem) according to the manufacturers' instructions. For quantitative real-time retro-transcriptase PCR (qPCR), specific probes from TaqMan[®] MicroRNA Assay (Applied Biosystem) were used. TaqMan[®] Universal Master Mix II, no UNG (Applied Biosystems) was used, in a Step One Plus machine (Applied Biosystems) according to the manufacturers' instructions. Relative levels of

expression were calculated using the comparative Ct method normalized to the internal control U6 snRNA.

3.6 Chromatin immunoprecipitation (ChIP) assay

Parental BT549 cells, BT549 cells transfected with SNAI1-MYC and NIH-3T3 cells were used for ChIP assays for each indicated experiment. First, cells were collected when at 80% confluency from 10 cm culture dish after trypsinization, and were resuspended in fresh culture medium. Then, paraformaldehyde (PFA) was added at a final concentration of 1%, followed by 10 min incubation at room temperature. Next, 1.25M glycine was added to a final concentration of 0.125M followed by 10 min incubation at room temperature. Medium was removed by centrifugation at 2000 rpm for 5 min, and cells were washed twice with cold phosphate-buffered saline (PBS). Then, cells were resuspended in 300 μ l of lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH8 and Protease inhibitor cocktail), followed by 15 min incubation on ice. The lysate was sonicated four times during 15 min at high voltage (H) with on/off intervals of 30 seconds each using Diagenode Bioruptor machine. Lysates were kept at 4°C throughout the whole procedure. Then the lysates were centrifuged at 13000 rpm for 10 min at room temperature. Supernatants (sonicated chromatin) were removed and diluted in 5 ml dilution solution (0.01% SDS, 1% Triton x-100, 2mM EDTA, 20 mM Tris-HCl pH8, 150 mM NaCl and Protease inhibitor cocktail). The diluted chromatin could be directly used or aliquoted and stored at -20°C. Aliquots were ran in agarose gel to check for proper fragmentation.

Antibodies (listed in [Table II](#)) were added to the chromatin dilutions and incubated overnight (O/N) at 4°C in a rotating rotor (1 μ l of antibody in 500 μ l of chromatin dilution). At the same time, Dynabeads protein A were incubated with blocking solution (lysis buffer, 0.05% BSA, 10 μ g/ml salmon sperm DNA and Protease inhibitor cocktail) O/N at 4°C on rotating rotor. Next, blocking solution was removed from the beads using magnetic concentrator, and the chromatin-antibody mixes were added to the beads, followed by incubation for 4 hours at 4°C in rotating rotor. Then, the beads were washed in four consecutive steps with wash buffers (WB) 1 to 4 using magnetic concentrator. (WB1: 0.1% SDS, 1% Triton X-100, 2mM Tris-HCl pH8 and 150 mM NaCl; WB2: 0.1% SDS, 1% Triton X-100, 2mM Tris-HCl pH8 and 500 mM NaCl;

WB3: 1mM EDTA, 10mM Tris-HCl pH8, 1% Igepal (NP40), 1% sodium deoxycholate and 0.25M LiCl; WB4: 10mM Tris-HCl pH8 and 1mM EDTA). Then 10% Chelex was added to the beads, followed by vortexing and incubation at 95°C for 10 min. Proteinase K solution was added followed by vortexing and incubating at 55°C for 30 min while shaking, and then the mixture incubated at 95°C for 10 min while shaking. Samples were centrifuged at 13000 rpm for 2 min and the supernatants, containing DNA, were used for qPCR assays.

3.7 Mouse, chicken and zebrafish embryo sections

C57BL/6J wild type mice were used. Embryos were staged as embryonic day according to days *post coitum*. Wild type Zebrafish strain AB were maintained at 28 °C under standard conditions, and the embryos were staged as previously described (Kimmel et al., 1995). Fertilized hen eggs were incubated in an humidified incubator at 37°C, and embryos were staged as previously described (Hamburger and Hamilton, 1951).

For *in situ* hybridization to detect mRNA (see protocol below), mouse and chicken embryos that were dehydrated through a series of increasing methanol:PBS-T concentration (25%, 50%, 75% and 100%) and twice in butanol, then embedded in paraffin O/N. Sectioning was performed in a Leica RM2245 microtome at 7µm thickness.

For *in situ* hybridization to detect mature miRNA, mouse embryos were fixed in 4% PFA-DEPC (Diethyl pyrocarbonate) O/N. The next day embryos were washed twice with PBS, 0.1% Tween 20 before being embedded in 15% sucrose. Upon sinking, embryos were embedded in 30% sucrose, and after sinking they were kept in fresh 30% sucrose O/N. Embryos were then kept in a 1:1 mix of 30% sucrose:OCT for 30 minutes while rolling, before embedding in OCT. Embedded embryos were kept on dry ice and transferred to -80°C before sectioning. OCT-embedded embryos were cryosectioned in a SLEE MNT cryotome at 10µm, dried for 2 hours at room temperature before either being directly used for *in situ* hybridization (see protocol below) or stored at -80°C.

Zebrafish embryos that were subjected to double-fluorescent *in situ* hybridization (see protocol below) were directly embedded in 4% low-melting agarose

and sectioned using a Leica VT1000S vibratome at 100 μ m, mounted using Dako fluorescent mounting medium and subjected to Confocal microscope imaging.

All animal procedures were conducted in compliance with the European Community Council Directive (2010/63/EU) and Spanish legislation. The protocols were approved by the CSIC Ethical Committee and the Animal Welfare Committee at the Institute of Neurosciences, Alicante.

3.8 RNA and miRNA *in-situ* hybridization

3.8.1 Probe synthesis

Digoxigenin (DIG)- or fluorescent (FLOU)-labeled probes were synthesized from mouse *Snail1* (Sefton et al., 1998), mouse *Prrx1*, chicken *SNAIL1* and *PRRX1*, and zebrafish *prrx1a/b* and *snail1a/b* (Ocana et al., 2017, Ocana et al., 2012) probe constructs. Constructs (1 μ g) linearized with proper restriction enzymes were used for RNA probe transcription, using Promega kit including 5x transcription buffer, 0.1M DTT, T7/T3/SP6 RNA polymerases, and placental ribonuclease inhibitor (50 U/ μ l), plus nucleotide mix (Roche). The mix was incubated at 37 °C for 3h. For purification of RNA probes, the products were mixed with 1/10 3M NaAc and 2.5volume of ethanol and incubated at -20 °C O/N. The product was then centrifuged 30 min at 4 °C, then washed twice with 70% ethanol, and dissolved in 1:1 volume nuclease free water and formamide. The quality of probes were checked by running in 1% agarose gels, and if satisfactory immediately used or stored at -80 °C.

3.8.2 Whole-mount mRNA visible and fluorescent *in situ* hybridization

Whole-mount *in situ* hybridization (ISH) was carried out as described previously (Nieto et al., 1996). Briefly, mouse, chicken and zebrafish embryos were fixed in 4%PFA-DEPC O/N. Zebrafish embryos were then dechorionated in cold 4%PFA-DEPC. Embryos then were washed in PBS 0.1% Tween20 (PBS-T) three times. All washes throughout the ISH were carried out on a rocker at the indicated temperature. Next, embryos were dehydrated through a series of increasing methanol concentration in PBS-T (25%, 50%, 75% and 100%) and then twice in methanol 100% and kept O/N at

-20 °C. Then, embryos were rehydrated through methanol:PBS-T in reverse order and washed twice in PBS-T at the end. For fluorescent ISH, embryos were incubated in 1% hydrogen peroxide (H₂O₂) for 10 min, and then washed three times with PBS-T. After that, embryos, depending on species and developmental stage, were treated with 10 µg/ml proteinase K in PBS-T between 3 to 6 min at room temperature. Then they were re-fixed with 4%PFA-DEPC for 20 min at RT, and washed twice. Embryos were then incubated with prehybridization solution (50% formamide, 5X SSC, 2% Boehringer blocking powder, 0.1% Tween 20, 50 µg/ml heparin, 1mg/ml t-RNA, 1mM EDTA, 0.1% CHAPS) at 60 °C for 30 min first, and then O/N upon refreshing prehybridization solution. Embryos were either used the next day for ISH or stored at -20 °C.

Prehybridized embryos were incubated with 1mg/ml of DIG or FLUO probes O/N at 60 °C. The next day, embryos were washed several times first with 2X SSC, 0.1% CHAPS and then 0.2X SSC, 0.1% CHAPS, and then with KTBT washing buffer (50mM Tris-HCl pH7.5, 150mM NaCl, 10mM KCl, 0.1% Triton X-100 in H₂O). After the washes, embryos were incubated in blocking solution (15% sheep serum, 0.7% Boehringer blocking solution, 0.1% Triton X-100 in KTBT) for 3h at 4 °C. Then embryos were incubated with anti-bodies (1/1000 anti-DIG-AP, 1/500 anti-DIG-POD or anti-FLUO-POD, depending on the experiment) in blocking solution O/N at 4 °C. The whole next day, embryos were washed many times in KTBT buffer and kept O/N in KTBT at 4 °C.

For visible ISH (chemical development of signal) embryos were washed in NTMT (100mM Tris-HCl pH9.5, 59mM MgCl₂, 100mM NaCl, 0.1% Tween-20, 1mM levamisole in H₂O) buffer three times prior to developing the signal. Embryos were then incubated with NTMT containing freshly-added 3 µl NBT and 2.6 µl BCIP per 1ml (developing solution), in the dark at RT until the color reaction develops. After obtaining the desired signal level in positive tissues, embryos were washed several times in KTBT and left O/N at 4 °C. After hybridization embryos were fixed in 4% PFA, washed in PBS and imaged in a Leica M125 dissecting scope with a Leica DFC 7000T digital camera (Leica, Wetzlar, Germany). Some embryos were embedded in paraffin, and sections were obtained at 7 µm. Sections were photographed under a Leica DMR microscope (Leica, Wetzlar, Germany).

For double fluorescent ISH, DIG and FLUO labeled probes were mixed and added to the hybridization solution. After the washing steps with SSC buffer and blocking, embryos were incubated with the first antibody (1/500 anti-DIG-POD or anti-FLUO-POD). The next day, after 2 times washing with KTBT, embryos were incubated with Amplification solution (TSA[®] fluorescein detection kit, PerkinElmer) for 2 min to adjust the pH. Cy3 (1/100) or FITC (1/50) were added to the Amplification solution for developing red or green colors, respectively. The embryos were then incubated in the mix for 45 min in dark at room temperature. Next, the samples were washed 5 times in KTBT, and incubated in 2% H₂O₂ for 2 hours, then washed 5 times in KTBT. Embryos were then incubated in blocking solution for 3 hours at 4 °C, prior to adding the second antibody (1/500 anti-DIG-POD or anti-FLUO-POD) in which they were incubated O/N at 4 °C. The next day, fluorescent developing procedures were repeated for the other color, plus stained with DAPI. After washing in KTBT, embryos were imaged with Olympus FV1200 confocal microscope and subjected to mosaic merge using Image J software, for pictures of whole-mounted embryos. Some embryos were embedded in 4% low-melting agarose and sectioned using a Leica VT1000S vibratome at 200µm and subjected to confocal microscope imaging.

3.8.3 Mature miRNA *in situ* hybridization

For mature miRNA *in situ* hybridization the commercially available mouse miR-322 (Mir-15-P1d), miR-503 (Mir-15-P2d), miR-15b (Mir-15-P1b), miR-16 (Mir-15-P2a/b) and scramble LNA probes, which were 5'-3' DIG labelled, were purchased from Exiqon. *In situ* hybridization was performed on 10 µm frozen embryo sections, using the EDC method, as previously described (Pena et al., 2009). Every step was performed in a wet chamber (50% formamide). Dried cryo-sections were treated with 20mg/ml proteinase K (pH7.4) in Tris-buffered saline (TBS) 1x (for 10x TBS, 69.6g Tris, 87.6g NaCl+800ml water, adjust pH to 7.6 and add water up to 1L) for 20 min at RT, then washed 2 times in TBS and fixed in 4% PFA for 10 min at room temperature. Then, sections were washed once with TBS 0.2% glycine for 5 min, and washed twice with TBS.

Samples were then incubated twice for 10 min in Imidazole 0.13M buffer (for 160ml of buffer, 1.6ml of 1-Methylimidazole was added to 130ml of water, pH was

adjusted to 8 with HCl, and then 16ml of NaCl 3M was added and then water to final volume). Sections were fixed in 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) solution (176µl of EDC (Sigma) was added to 10ml of Imidazole buffer, then pH was re-adjusted to 8 by adding HCl) for 3 hours at room temperature. Samples were washed once with TBS 0.2% glycine, and twice with TBS. Next, freshly prepared 0.1M Triethanolamine (TEA), 0.5% acetic anhydride was added for 30min at room temperature. After washing twice with TBS, the sections were prehybridized with hybrid-mix solution (50% formamide, 5x SSC, 5x Denhardt's solution (Applichem), 250 µg/ml yeast tRNA (Sigma), 500 µg/ml salmon sperm DNA (Sigma), 2% (w/v) Blocking Reagent (Roche), 0.1% 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPs) (Sigma), 0.5% Tween) for 2 hours at room temperature.

For hybridization, 4 pmol of DIG-labeled LNA probes were diluted in of hybrid-mix solution for each slide, and covered with Parafilm M (Sigma). The slides were kept in a sealed humidified chamber for at least 16 hours at a temperature 20°C below the melting temperature of the miRNA-LNA probes. Next day, sections were washed twice for 30min in washing solution (50% formamide, 1X SSC, 0.1% Tween 20) at hybridization temperature, and then washed once with 0.2% SSC for 15 min at room temperature plus one wash with TBS 0.1% Tween 20. Slides were then incubated with 3% hydrogen peroxide, 0.1% Tween20 in TBS for 30 min, and then washed 3 times in TBS 0.1% Tween 20 at room temperature. Subsequently, sections were blocked by adding 0.5% Blocking Reagent (Roche), 10% sheep serum, 0.1% Tween 20 for 1 hour at room temperature. Then, samples were incubated with anti-DIG antibody in blocking solution O/N at 4 °C. The next day, sections were washed 5 times with TBS 0.1% Tween 20, and incubated in NTMT buffer 3 times. Slides were next immersed in chambers containing developing solution (detail in RNA ISH) at 37 °C until the color reaction developed. The sections were then washed several times in TBS and fixed in PFA before being photographed under a Leica DMR microscope (Leica, Wetzlar, Germany).

3.9 Western blot

E11.5 mouse embryos were dissected and lysed directly in RIPA buffer (25mM Tris-HCl pH 7.6, 150mM NaCl, 1% NP40, 1% Sodium Deoxycholate, 1% SDS and

Proteinase inhibitor cocktail). Homozygous LacZ-Knock-in *Prrx1* mutant embryos were identified among littermates by PCR genotyping and used for further analyses, together with wild type embryos. Lysates were then homogenized followed by heating, sonication (6 rounds of 15 min, 30 sec ON/30sec OFF, high). Samples were then loaded and subjected Coomassie Blue staining to determine the quality and concentration of the samples. Next, protein samples were loaded and migrated in 12% acrylamide gel and then transferred to PVDF western blotting membrane (Roche). The membranes were blocked with 10% milk and incubated with anti-Snail1 antibody overnight (Rabbit polyclonal, Cell Signaling). The next day, membranes were extensively washed and incubated with secondary Peroxidase goat anti-Rabbit antibody, washed and developed using Chemiluminescent HRP Substrate (Immobilon™ Western, Millipore). Pictures were obtained using Amersham Imager 680 (GE Healthcare). β -actin (Rabbit polyclonal, Abcam) was used as housekeeping protein for normalization.

3.10 Immunofluorescent (IF) staining

Whole-mount mouse embryos were fixed in 4% PFA for 2 hours at 4°C. Then, embryos were dehydrated in a series of PBS 1% Triton x100 (PBS-T):methanol proportions (25%, 50%, 75% and 100%), then kept in 100% methanol O/N at -20 °C. Embryos were then rehydrated in the reverse order of PBS-T:methanol proportions. Antigen retrieval was performed by treating the mouse embryos with either 150 mM Tris-EDTA pH 9.0 (for *Prrx1*) or 10mM Sodium Citrate pH 6.0 (for Snail1)-1% Triton x100 buffer at 70 °C for 20 minutes, following three washes with PBS-T. Embryos were blocked with 5% NGS 1% BSA 1% Tx-100 for 5 hours or O/N at 4°C and then incubated with the primary antibodies 48 hours at 4°C. After several washes for several hours in PBS-T, the embryos were incubated with the secondary antibodies and DAPI O/N at 4°C. After washing the secondary antibody with PBS-T, embryos were subjected to imaging. Pictures of whole-mount embryos were taken with an Olympus FV1200 confocal microscope with 20x (E8.5) or 10x (E9.5) objective and then subjected to mosaic merge using Image J software.

For cell lines IF, cells were cultured and treated on cover-slips in 6-well plates and collected at corresponding time points, fixed with 4% PFA for 15 min at room temperature, and washed three times with PBS. After washing fixed cells were either

directly subjected to IF or stored at 4°C. Cells were blocked with 5% NGS 1% BSA 0.2% Triton x-100 1 hour at room temperature and incubated with the primary antibodies O/N at 4°C. After washing three times with PBS, cells were incubated with the secondary antibodies and DAPI 1 hour at room temperature. After washing the secondary antibody with PBS, cells were photographed. Pictures were taken with Leica DMI8 microscope and HAMAMTSU C11440 digital camera.

3.11 In silico analyses

3.11.1 Single cell RNA-seq public data

Processed datasets of publically available data for single cell RNA-seq were downloaded from NCBI Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Data included single cell RNA-seq from developing mouse embryos, different cancer types and fibroblasts from pulmonary fibrosis (Table III). Among analyzed single cells, the ones with no value for both *Snail1* and *Prrx1* were excluded. Hierarchical clustering was performed using GENE-E (version 3.0.215, Broad Institute, Inc.). Values were subjected to correlation analyses using Prism (GraphPad Softwares, Version 6.01, 2012) calculated based on Spearman r.

3.11.2 Meta-analysis of oncogenomic data

To assess the putative correlations between the expression of *SNAIL1*, *PRRX1* and mature *miR-15* family members with distant metastasis-free survival in all subtypes ($n=3951$) or lymph-node positive patients ($n=1133$), all available data sets (GSE6532, GSE20711, GSE7390, GSE21653, GSE31519, GSE5327, GSE17907, E-MTAB-365, GSE37946, GSE2034, GSE2990, GSE17705, GSE1456, GSE12093, GSE9195, GSE45255, GSE20685, GSE12276, GSE2603, GSE16391, GSE42568, GSE11121, GSE3494, GSE16446, GSE4611, GSE26971 and GSE19615) were analyzed and Kaplan–Meier plots were generated using <http://kmplot.com> (Lanczky et al., 2016, Györfy et al., 2010). The patient samples were grouped as either high or low for the expression of the genes of interest, and the auto-select best cut-off was chosen for computation over the entire data set.

3.12 Statistical and data analysis

Images were prepared using Adobe Photoshop and Adobe Illustrator from CS6. Sample size was estimated using GPower 3.1, and values were set at $p = 0.05$ and $\beta = 0.8$. All statistical analyses were performed using Microsoft Excel 2013 and Prism (GraphPad Softwares, Version 6.01, 2012). For reporter assays and qRT-PCR experiments, the corresponding treatments were compared with controls using Student's t-test or One-way ANOVA with Bonferroni's multiple comparison test. Spearman r was used for correlation test of single cell RNA-seq data. All bar graphs represent mean + SEM (Standard Error of the Mean). Statistical significances were as follows: * = $P \leq 0.05$, ** = $P \leq 0.01$ and *** = $P \leq 0.001$.

3.13 Tables

Table I. Primer sequences	
Human miR424-503-promoter-F (Cloning)	GACTGCTAGCTTCAATATCATTCTCACAAATACAAAATAG
Human miR424-503-promoter-R (Cloning)	GATCCTCGAGTGGAACAACGTAGTGGGTGA
Human miR424-503-F (Cloning)	GACTGGATCCGGCTTCCTTCAGTCATCCAG
Human miR424-503-R (Cloning)	GACTCTCGAGTCTACCTGAGCAGGGAAAGG
h-424-mut-F	GATCCCCCTTCATTGACTCCGAGGGGATACTATTATAATTCATGTTTTGAA GTGTTCTAAATGGTTC
h-424-mut-R	GAACCATTTAGAACACTTCAAACATGAATTATAATAGTATCCCCTCGGAG TCAATGAAGGGGGATC
h-503-mut-F	GTGCCCCGCGCTCAGCCGTGCCCTTATTATGGGAACAGTTCTGCAGTGA
h-503-mut-R	TCACTGCAGAACTGTTCCCATATAAAGGGCACGGCTGAGCGCGGGCAC
miR-15b-16-2-cloning-F	GACTGGATCCACTAAAGCTTGAAAGAGTGTTCCTCTGT
miR-15b-16-2-cloning-R	GACTCTCGAGATAAACAAAAGGGACAGATTATCAAAAAG
QPCR HsPrrx1 FW	CTGATGCTTTTGTGCGAGAA
QPCR HsPrrx1 RV	ACTTGGCTCTTCGGTTCTGA
hSNAI1-F	GCTGCAGGACTCTAATCCAGA
hSNAI1-R	ATCTCCGGAGGTGGGATG
QPCR HsPrrx1 FW	CTGATGCTTTTGTGCGAGAA
QPCR HsPrrx1 RV	ACTTGGCTCTTCGGTTCTGA

premir424-F	TTGACTCCGAGGGGATACAG
premir424-R	GACCCACCTTCTACCTTCC
premir503-F	GCGAGTCGAGGAGAGACG
premir503-R	GAACGGCAGTCCCAGACTTA
cfSnail-Ex1-F2	AACTGCAAATACTGCAACAAGGAATAC
cfSnail-Ex2-R2	CAGGAAAACGGCTTCTCACCG
cfPrrx1-Ex2-F	CGCGTCTTTGAGAGAACACAC
cfPrrx1-Ex3-R	GCATGGCTCTCTCATTCTTC
cfRS17-F	CAAGATCGCAGGCTATGTGA
cfRS17-R	CCTCGATGATCTCCTGATCC
hsa-TBP-F	CGGCTGTTTAACTTCGCTTC
hsa-TBP-R	CACACGCCAAGAAACAGTGA
h-PRRX1-del1-F	TTATTCGTCTACCTTCAGAAGATTCTCTTCCACACTAATTGG
h-PRRX1-del1-F	CGAATTAGTGTGGAAGAGAATCTTCTGAAGGTAGACGAATAA
h-PRRX1-del2-F	TTCAGAGTGGGTTTTTTTTTAAATGCAATATTGTATTCAAACAAAAAGAGG G
h-PRRX1-del2-F	CCCTCTTTTTGTTTGAATACAATATTGCATTAACAAAAAACCCACTCTGAA
h-PRRX1-del3-F	GGAGAAAAGTTGTGGAGTTTACATGAACATTTGCTAAACATGTTTT
h-PRRX1-del3-F	AAAACATGTTTAGCAAATGTTTCATGTAAACTCCACAACCTTCTCC
miR-15b-16-2-cloning-F	GACTGGATCCACTAAAGCTTGAAAGAGTGTTCTTCTGT
miR-15b-16-2-cloning-R	GACTCTCGAGATAAACAAAAGGGACAGATTATCAAAAAG
mmu-miR,15bmut-F	GCTGAGTCCTGTCTTTTGAACCTTAAAGTACTGTTCTGTCGACATCATGGT TTACATACTACAG
mmu-miR,15bmut-R	CTGTAGTATGTAAACCATGATGTGCGACGAACAGTACTTTAAGGTTCCAAAA GACAGGACTCAGC
mmu-miR,16mut-F	GTATTATGTTTGGATATCTGACATGCTTGTTCCACTCTTCGTGCGACGTAAA TATTGGCGTAGTGAAATAAAT
mmu-miR,16mut-R	ATTTATTTCACTACGCCAATATTTACGTGCGACGAAGAGTGGAACAAGCATG TCAGATATCCAAACATAATAC
m497-PxB-F	CTAAAAACAGGCCTTATTGACTACA
m497-PxB-R	GATGTGGCTTTACTTTGTGAAGATT
m497-up-F	AGGATATTAGAATTGGGGAGGTTAGT
m497-up-R	TGAGTCCAATACATTAGAAAGAGGAGT
m497-dw-F	ACTTGCTCGTACAGGTTGTATATGTTCT
m497-dw-R	CTTGGGAAGACTAATAGGAGTTAAAGTG
miR-195mut-F	GTTGCCACACCCAACTCTCCTGGCTCTTCGTGCGACAGAAATATTGGCAT GG
miR-195mut-R	CCATGCCAATATTTCTGTGCGACGAAGAGCCAGGAGAGTTGGGTGTGGGC AAC
miR-497mut-F	CAGTCCTGCCCCCGCCCTCGTGCACACTGTGGTTTGTACGG
miR-497mut-F	CCGTACAAACCACAGTGTGCGACGAGGGGCGGGGGCAGGACTG
m195-497-clon-F2	GACTGGATCCCTAAACTACTTTTGCTGGTTCCTGATT
m195-497-clon-R2	GACTCTCGAGGACTTCTGTGTGATGGACATTTTATAC

SNAI1,3',MRE503-F	TTTGTATCCAGAGCTGTTTGGATACACGACGATTGAGCTACAGGACAAAG GCTGACAG
SNAI1,3',MRE503-F	CTGTCAGCCTTTGTCCTGTAGCTCAATCGTCGTGTATCCAAACAGCTCTG GATACAAA
h503P-del1-F	AAAAAAAAACATGTTTAGCAAATGTATCATGTAAACTCCACAACCTTTCTCC
h503P-del1-R	GGAGAAAGTTGTGGAGTTTTACATGATACATTTGCTAAACATGTTTTTTTTT
h503P-del2-F	CCCTCTTTTTGTTTGAATACAATATTAGCATTAAAAAAAAAACCCACTCTGA
h503P-del2-R	TCAGAGTGGGTTTTTTTTTAAATGCTAATATTGTATTCAAACAAAAAGAGG G
hPRRX1-P-F2	ATTAGTAAATGTGTGCGGTGTTTTCTTG
hPRRX1-P-R3	CTTTGTAGGGAATACAAGAAAATTGGAG
hPRRX1-P-R4	ATTAGTAAATGTGTGCGGTGTTTTCTTG
497P-PXB1-F	AGGAGGGAGTGACTTCCAAAA
497P-PXB1-R	AAATTTGGGGTCCTCAGATACC
497P-PXB2-F	CTACCCAGATGTCTTTGGAGGT
497P-PXB2-R	CCGAAACAAAATATGAGGGTGT
497P-dw-F	TTATTGAGATACGGGACACAGC
497P-dw-R	CTCCAGCCCCCTCCTCTATTTA
SMC4P-dw-F	AGCAAGACCCCATCTCTACAAA
SMC4P-dw-R	AGTGCACACGTAAAAGGACTGA
SMC4i-PXB1-2-F	TTGGATATGGTGGGAATCTTGT
SMC4i-PXB1-2-R	TCAATTGAACATGCACACAAAC
SMC4P-PXB3-F	GAACCTATTTCTTCCTGTGGGGTA
SMC4P-PXB3-R	CCTGGAAAAGACTGGGTACAG
PRRX1P-dw-F	GGACTTACAGACATTCCCTTGG
PRRX1P-dw-R	GGCCTGAGAAACAAATAGATGG
PRRX1P-PXB1-F	CTCCCTTTCTCTCTAACTCTGATGTTG
PRRX1P-PXB1-R	GCCAACATCAGAGTTAGAGAGAAAG
PRRX1P-PXB2-F	TTGATTGCCTGCATTCTTACTT
PRRX1P-PXB2-R	TCAATGAGTCCAAAATGTAAACC
PRXP-B1-F2	GTACTTTTCTCATTGCCCTAGTCCT
PRXP-B1-R2	GTAAGCATATTAAGGCTATTTTTGGTTC
PRXP-B1-F3	AAGAAGGAGATTGTGATGGAGAAAG
PRXP-B1-R3	TTTTCTTCTCAGTTGGATCAAGAA
PRXP-up-F	CCTAGCACCTGAGAAACATACTGATAAC
PRXP-up-R	ACTTAGCCTATAAATCAGAGAATACTTGT
PRXP-PXB3-4-F	AGACTATAACAGAGCAGATCAATTACCA
PRXP-PXB3-4-R	TCCCATATTTAATGTTTTAGGAGTCT
PRXP-PXB5-6-F	TTTATGAGGTGAACGCATTATCAG
PRXP-PXB5-6-R	TCTTGCTTGTTAAGTCTTTGTGG
SNAI-B-1-2-R	CAGTTAGTAGTCCAGGTGTCTTGC
PRRX1P-F01Xho	GACTCTCGAGCCTTTATTCTCTTTCTTAATCCTCTCAA
PRX-B1-R2Nco	GACTCCATGGGTAAGCATATTAAGGCTATTTTTGGTTC

PRXP-CpG-R	GACTCCATGGAGGTTTCACCCTTAAAGGACATTC
PRXP-COM-R	GACTCCATGGAGGTGACTGACGGAGAAGTTCTTT
m503-F2	GACTGCTAGCCCCTCCCAAATTATTCCAAA
SNAI1-PXB3-5-F	CAACAGTTCTTTACTTTGCTTGGTT
SNAI1-PXB3-5-R	ACAGGGAAGGATTAACACCTAAG
SNAI1-PXB6-F	GACTTTGGCTTTTACTCTGAGACAG
SNAI1-PXB6-R	ACAGAGGCAGTAAGCAGTCATTAAG
SNAI1-PXB7-F	TCCCTTTGCATTGTAATTATCTGTT
SNAI1-PXB7-R	GTTCCCTCCCTTATCCAGTGTTCAC
SNAI1-dw-F	GCCCAAATTGTCAGTTTCATAAATA
SNAI1-dw-R	ACAAAGATGTAAACCAAGATCTCCA
EGFP_fwd	CCGACCTCTCTCCCCAGGGGATCCGTGAGCAAGGGCGAGGAG
EGFP_rev	TAGTAGCTCCGCTTCCCTTGTACAGCTCGTCCATGC
PuroR_fwd	GCTGTACAAGGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCT GGAGACGTGGAGGAGAACCCTGGACCTATGACCGAGTACAAGCCCACG G
PuroR_rev	GTCATTGGTCTTAAAGGTACCTCAGGCACCGGGCTTGCG
hUbC_promoter-F0	TCTTTCCAGAGAGCGGAACA
YFP-BamHI-F	TTAAGGATCCGCTCGTTTAGTGAACCGTCAG
YFP-KpnI+stop-R	TTAAGGTACCTACTTGTACAGCTCGTCCATGC
hPGK promoter-F	GTAGTGTGGGCCCTGTTCTT
3'LTR-R	TCGTTGGGAGTGAATTAGCC

Table II, Antibodies	
Prrx1 (for ChIP)	Sigma, Rabbit polyclonal (HPA051084)
IgG	Diagenode, Rabbit (C15410206)
Myc tag	Abcam, goat pAb (ab9132)
Prrx1 (for IF)	Tanaka lab (Ocana et al., 2017)
Snail1	Cell Signaling, rabbit monoclonal (C15D3)
GFP	Chicken polyclonal, Aveslab (2BScientific), GFP-1020
DIG-AP	Fab fragments, sheep polyclonal, Roche (11093274910)
DIG-POD	Fab fragments, sheep polyclonal, Roche (11207733910)
FLUO-POD	Fab fragments, sheep polyclonal, Roche (11426346910)
β-actin	Rabbit polyclonal, Abcam

Table III, GEO datasets									
GEO acc. #	Sample scRNA-seq	Total cell #	# of cells	# of <i>Snail1</i> single + cells	# of <i>Prrx1</i> single + cells	# of cells <i>Snail1</i> _{high} and <i>Prrx1</i> _{low/-}	# of cells <i>Snail1</i> _{low/-} and <i>Prrx1</i> _{high}	# of cells <i>Snail1</i> _{high} and <i>Prrx1</i> _{high}	Ref.
GSE87038	E9.5 mouse embryos	764	440	219 (49%)	87 (19%)	245 (55%)	143 (32%)	54 (12%)	(Dong et al., 2018)
GSE87038	Somites, E9.5 mouse embryos	183	81	36 (44%)	10 (12%)	43 (53%)	26 (32%)	12 (14%)	(Dong et al., 2018)
GSE103322	18 head and neck cancer patients	5902	1718	480 (28%)	1005 (58%)	564 (32%)	1031 (60%)	123 (7%)	(Puram et al., 2017)
GSE75688	11 breast cancer patients	551	138	96 (69%)	22 (16%)	103 (74%)	28 (20%)	7 (5%)	(Chung et al., 2017)
GSE112294	18hpf zebrafish embryos (GSM3067194)	6962	1620	snail1a : 582 (36%) snail1b : 98 (6%)	prrx1a : 376 (23%) prrx1b : 171 (10%)	snail1a : 787 (48%) snail1b : 116 (7%)	prrx1a : 627 (38%) prrx1b : 332 (20%)	prrx1a : 52 (1.6%) prrx1b : 2 (0.1%)	(Wagner et al., 2018)
GSE104154	Fibroblasts in mouse pulmonary fibrosis	10410	1131	209 (18%)	831 (73%)	293 (26%)	836 (74%)	2 (0.1%)	(Xie et al., 2018)

Chapter 4

RESULTS

4.1 Prrx1 and Snail1 are expressed in complementary patterns in development and disease

4.1.1 Prrx1 and Snail1 complementary expression pattern is conserved in chicken and zebrafish embryos

As shown previously (Ocana et al., 2012), RNA *in situ* hybridization (ISH) using specific probes for *SNAIL1* and *PRRX1* shows that these two EMT-TFs are expressed in a complementary pattern during the development of the chicken embryo (Figure 12a-b). We have now analyzed the publicly available single-cell RNA-sequencing (scRNA-seq) data from developing chicken embryo at stage HH4 (GSE89910) and again, we find a complementary expression pattern between *SNAIL1* and *PRRX1* in the majority of cells analyzed, resulting in a significant negative correlation (Figure 12c). We then characterized the expression of the two EMT-TFs in zebrafish embryos, which bear two paralogs for each gene (*prrx1a*, *prrx1b* and *snail1a* and *snail1b*) due to the extra duplication in the teleost genome (Jaillon et al., 2004). We also found a complementary expression pattern. This was evident in the developing somites, where *snail* genes are abundantly expressed while *prrx1* genes expression is restricted to small cell populations (Figure 12d, arrows). Both *snail1* and *prrx1* genes are expressed in the cranial neural crest (Figure 12d), and transverse sections of double fluorescent ISH showed that they are also expressed in a complementary manner (Figure 12e).

We analyzed the publicly available scRNA-seq data from developing zebrafish embryos 18 and 24 hours post fertilization (hpf) (GSM3067194). Again, we found a complementary expression between *snail1a/b* and *prrx1a/b* in the majority of cells analyzed at both developmental time-points. We used Spearman's rank-order correlation test to measure the strength and direction of association between the two *snail1* and *prrx1* gene pairs, finding a significant negative correlation between them (Figure 12f-i). Altogether, these findings confirm that the complementary expression of *SNAIL1* and *PRRX1* previously found in the chick is conserved in zebrafish embryos.

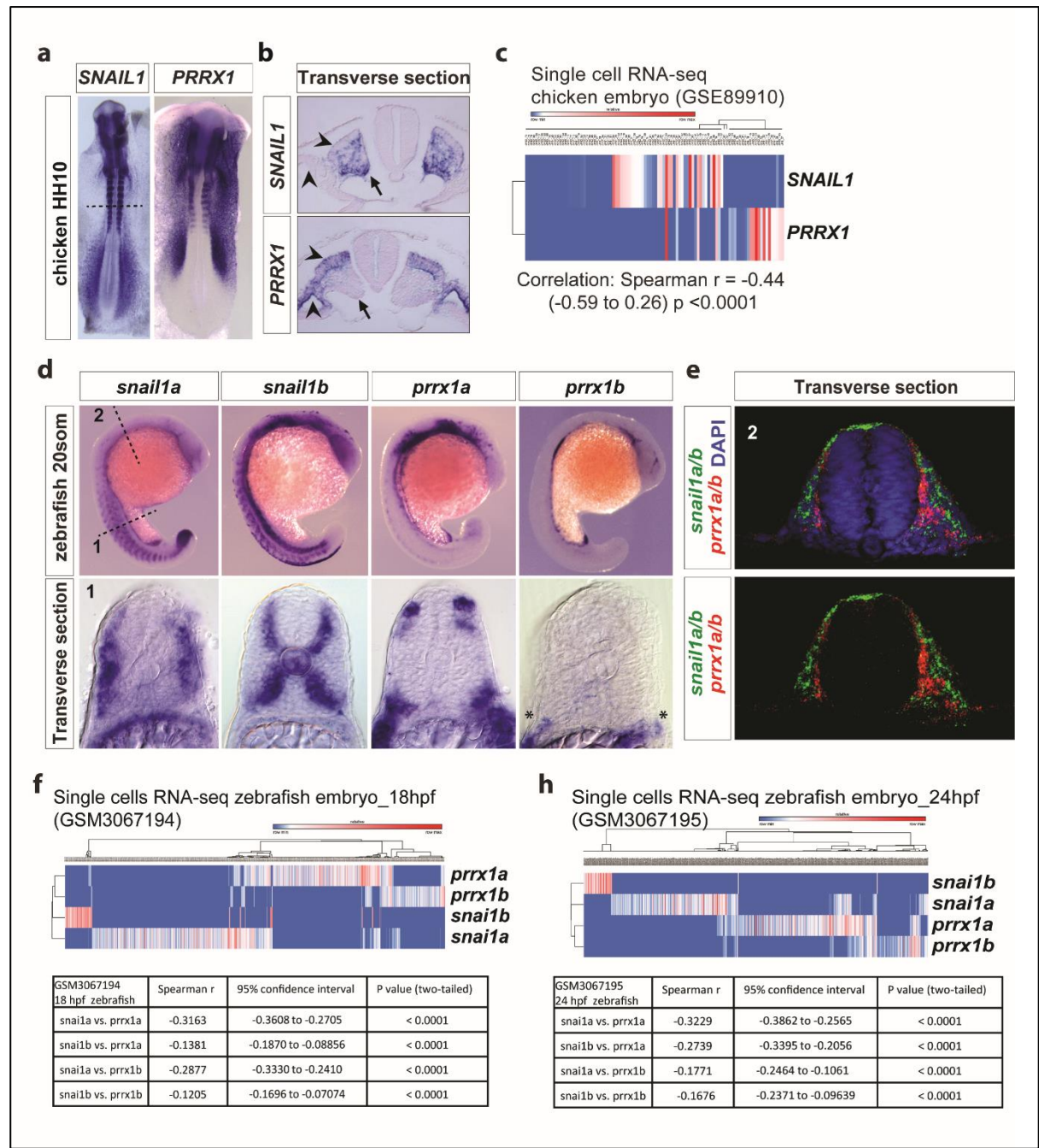


Figure 12. Prrx1 and Snail1 complementary expression pattern is conserved in chicken and zebrafish embryos. (a) Dorsal view of HH10 chicken embryos showing *PRRX1* and *SNAIL1* expression. (b) Transverse sections of chicken embryos showing the complementary expression pattern of *SNAIL1* and *PRRX1*. Within the somites, *SNAIL1* is expressed in the sclerotome (arrows) and *PRRX1* in the dermomyotome (arrowheads). (c) Hierarchical clustering heatmap of scRNA-seq data from stage HH4 chicken embryos, from the public database GEO: GSE89910. It shows a trend of *SNAIL1* and *PRRX1* expression being mutually exclusive in the majority of cells, with a significant negative correlation. (d) Lateral view of 20-somite zebrafish embryos showing *snai1a*, *snai1b*, *prrx1a* and *prrx1b* expression in whole-

mount (top) and transverse sections (down), again showing complementary patterns in somites. (e) Transverse section of zebrafish embryo in the cranial neural crest region in the showing complementary expression of *snail1* (green) and *prrx1* (red) genes together in the same embryo. (f, g) Hierarchical clustering heatmaps of scRNA-seq data from 18 and 24hpf zebrafish embryos, from public database GEO: GSE112294, (GSM3067194 and GSM3067195, respectively), showing *snail1a/b* and *prrx1a/b* transcription being mutually exclusive in the majority of cells. Interestingly, the 4 genes seem to be expressed in different cells, with significant negative correlations between any of the gene pairs.

Arrowhead: sclerotome; arrow: dermomyotome; asterisk: *prrx1b* expression; RNA-seq: RNA sequencing; vs: versus; hpf: hours post fertilization.

4.1.2 Prrx1 and Snail1 complementary expression pattern is conserved in mouse embryos

To assess whether this complementarity was also conserved in mammals, we characterized the expression of *Snail1* and *Prrx1* in mouse embryos. At E8.5, embryos manifest a complementary expression pattern for *Snail1* and *Prrx1* at whole-mount level, particularly evident in the mesodermal populations along the medio-lateral axis and in the somites (Figure 13a). Transverse sections of embryos at this stage confirms this complementarity in the neural crest populations, where *Snail1* is expressed in the premigratory neural crest (PNC, arrows) and both are expressed in different migratory neural crest populations (MNC, arrowheads), (Figure 13b, 1) that better resolves at E9.5 (see below). In more posterior tissues, the complementarity is evident in the mesodermal cells, where *Snail1* is highly expressed in the pre-somitic mesoderm (PSM, arrow) while *Prrx1* is expressed in the lateral plate mesoderm (LPM, arrowhead) (Figure 13b, 2 and 13c). Expression is also complementary in the somites and as in the chick, *Snail1* is highly expressed in immature somites and in the sclerotome (SC) and *Prrx1* in the dermomyotome (DM). Also, LPM cells are positive for *Prrx1* expression but not *Snail1*, (Figure 13c). Analysis of transverse sections at the level of the branchial arches, also showed complementarity, where in more internal cell populations *Snail1* was highly expressed concomitant with low *Prrx1* expression (inset box 3), while more ventral cell populations showed high expression of *Prrx1* and lower *Snail1* levels (inset box 4) (Figure 13d). A similarly complementary was also evident in the mature somites.

And as in the chick, *Snail1* is highly expressed in the SC (inset box 4') and *Prrx1* in the DM (inset box 3') (Figure 13e).

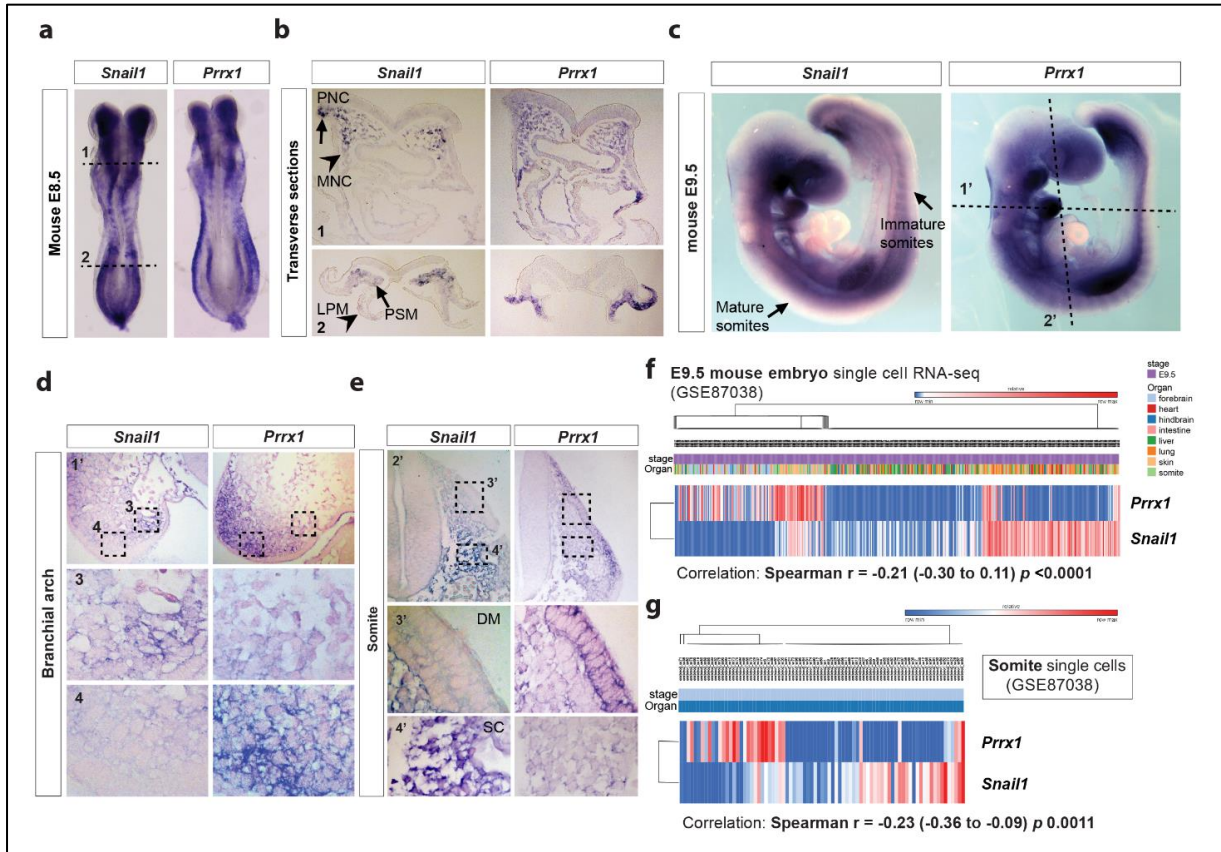


Figure 13. *Snail1* and *Prrx1* have complementary expression in the developing mouse embryo. (a) Expression of *Snail1* and *Prrx1* in dorsal views of E8.5 mouse embryos. (b) Transverse sections of E8.5 embryos from the regions marked by dashed lines (1 and 2) in (a), showing complementary expression of *Snail1* and *Prrx1* in PNC (1, arrow), MNC (1, arrowhead), and mesodermal populations (PSM, 2, arrow; LPM, 2, arrowhead). (c) Lateral view of E9.5 mouse embryos. (d, e) Transverse sections of the embryos shown in (c) to better observe the branchial arches (1') and the somites (2'), together with higher power pictures of the boxed areas (3, 3', 4, 4'). Complementary expression is observed in the branchial arches and the somites. (f) Hierarchical clustering heatmaps of scRNA-seq data from E9.5 mouse embryos for *Snail1* and *Prrx1* expression (GEO: GSE87038) showing a significant negative correlation. (g) A similar analysis obtained from E9.5 mouse embryo somites (GEO: GSE87038), again showing a significant negative correlation between *Snail1* and *Prrx1* expression.

PNC: premigratory neural crest; MNC: migratory neural crest; LPM: lateral plate mesoderm; PSM: presomitic mesoderm; DM: dermomyotome; SC: sclerotome

The complementary expression of *Snail1* and *Prrx1* in the E9.5 mouse embryo sections is supported at whole organism level by single cell RNA sequencing (scRNA-seq). Analysis of publicly available scRNA-seq (GSE87038) clearly shows this complementary expression at single-cell level, which results in a statistically significant negative correlation (Figure 13f). Interestingly, there are cells that express both transcripts suggesting a transitory state as observed in tissue sections residing between high *Snail1* and high *Prrx1* populations (Figures 13b and f). We also looked at single cell analyses obtained from the somitic region (GSE87038) and similarly, we observed a clear mutually exclusive expression pattern and a negative correlation of *Snail* and *Prrx1*, and some cells expressing both TFs (Figure 13g). These data confirm that complementary expression pattern of *Snail1* and *Prrx1* is also observed in the mouse, suggesting a conserved dynamic mutually exclusive expression code for these two EMT-TFs in cells undergoing EMT during development.

4.1.3 *Prrx1* and *Snail1* expression pattern is mutually exclusive in cancer and fibrosis

After finding the complementary/mutually exclusive expression pattern of *Snail1* and *Prrx1* during vertebrate development, we examined whether this holds true in pathology. Therefore, we first looked at the expression levels of *SNAIL1* and *PRRX1* in human breast cancer cell lines. We chose SUM149PT cells as they have high *SNAIL1* and low *PRRX1* expression; MDA436 cells with intermediate levels of both, and BT549 cells with high *PRRX1* and low *SNAIL1* expression. Immunofluorescence analyses (IF) confirmed previous results obtained in microarrays of (Klijn et al., 2015) (Figure 14a). Interestingly, even in cells with high *SNAIL1* or *PRRX1* protein levels, a complementarity was observed at single cell resolution (Figure 14a). Signal intensity analyses for *SNAIL1* and *PRRX1* the three cell lines show a trend for a negative correlation at protein level, with gradual reduction of *SNAIL1* concomitant with a gradual increase for *PRRX1* levels from SUM to BT cells (Figure 14b). Interestingly, MDA436 cells harbor cells expressing both *SNAIL1* and *PRRX1*, however, with intermediate levels for both (Figure 14b). This suggests that *SNAIL1* or *PRRX1* at levels cannot be co-express together, and cells can be found to express both only when they are expressed at low levels.

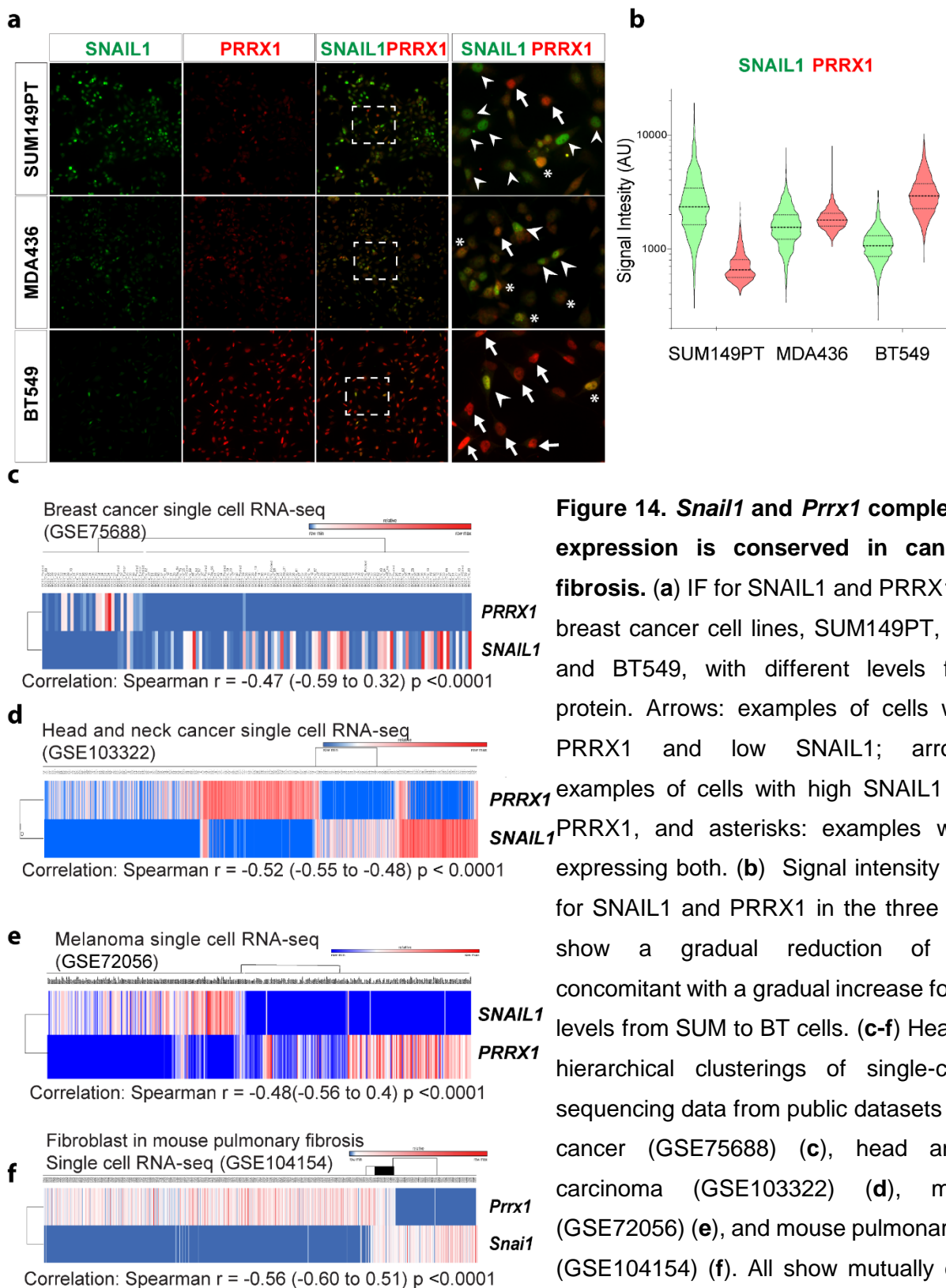


Figure 14. *Snail1* and *Prrx1* complementary expression is conserved in cancer and fibrosis. (a) IF for SNAIL1 and PRRX1 in three breast cancer cell lines, SUM149PT, MDA436 and BT549, with different levels for each protein. Arrows: examples of cells with high PRRX1 and low SNAIL1; arrowheads: examples of cells with high SNAIL1 and low PRRX1, and asterisks: examples with cells expressing both. (b) Signal intensity analyses for SNAIL1 and PRRX1 in the three cell lines show a gradual reduction of SNAIL1 concomitant with a gradual increase for PRRX1 levels from SUM to BT cells. (c-f) Heatmaps of hierarchical clusterings of single-cell RNA sequencing data from public datasets of breast cancer (GSE75688) (c), head and neck carcinoma (GSE103322) (d), melanoma (GSE72056) (e), and mouse pulmonary fibrosis (GSE104154) (f). All show mutually exclusive expression of *Snail1* and *Prrx1* with a significant negative correlation using Spearman r correlation analysis.

Next, we examined samples from mice and patients with diseases in which EMT has been shown to play essential roles (Nieto et al., 2016, Thiery et al., 2009, Nieto, 2013). We took again advantage of the available scRNA-seq studies in different human cancer types and mouse pulmonary fibrosis, and found that in breast cancer (GEO: GSE75688) (Chung et al., 2017) *SNAIL1* and *PRRX1* depict a mutually exclusive expression with a significant negative correlation (Figure 14c). In addition, in larger sRNA-seq dataset from head and neck carcinoma patients (GEO: GSE103322) (Puram et al., 2017), melanoma patients (GEO: GSE72056) (Tirosh et al., 2016) and single fibroblast cells from mouse lungs with pulmonary fibrosis (GEO:GSE104154) (Xie et al., 2018), *Snail1* and *Prrx1* are also expressed in a complementary manner and negatively correlated, with some small cell populations expressing both (Figures 14d-f).

Taken all together, these findings strongly suggest that there is a fundamental cross-talk regulatory mechanism between these two EMT-TFs that is conserved not only during vertebrate development but also across different pathological EMT, such as those occurring in cancer and fibrosis.

4.2 *Snail1* directly represses *Prrx1* transcription, while *Prrx1* induces its own expression

We next wanted to understand the molecular basis of the mutually exclusive expression of *Snail1* and *Prrx1*. Considering that they are transcription factors (TF) and that during embryonic development there are many examples of mutually regulated TF pairs (See for instance (Acloque et al., 2011, Acloque et al., 2012), we set up to study this possibility for *Snail1* and *Prrx1*.

As *Snail1* has been described as a potent transcriptional repressor, we examined whether it can directly target *Prrx1* transcription. We looked for consensus *SNAIL1* E-boxes (Villarejo et al., 2014) within the mouse *Prrx1* promoter, and found several predicted binding sites across 2.5 kb upstream of the transcription starting site (TSS). We performed ChIP assay in NIH3T3 cells transiently overexpressing *Snail1*, and found at least two different regions with significant enrichment for *Snail1* protein occupancy; binding site (BS)3: -2100 and BS4: -265 (Figure 15a). We then analyzed

the human *PRRX1* promoter (8 kb upstream of *PRRX1* TSS) and we also found several putative SNAIL1 E-boxes. ChIP assays in BT549 cells transiently overexpressing SNAIL1 also show a tendency for enrichment of SNAIL1 binding in three different regions; BS3: -5631, BS4: -4102 and BS5: -1689, with BS4 being significantly enriched (Figure 15b).

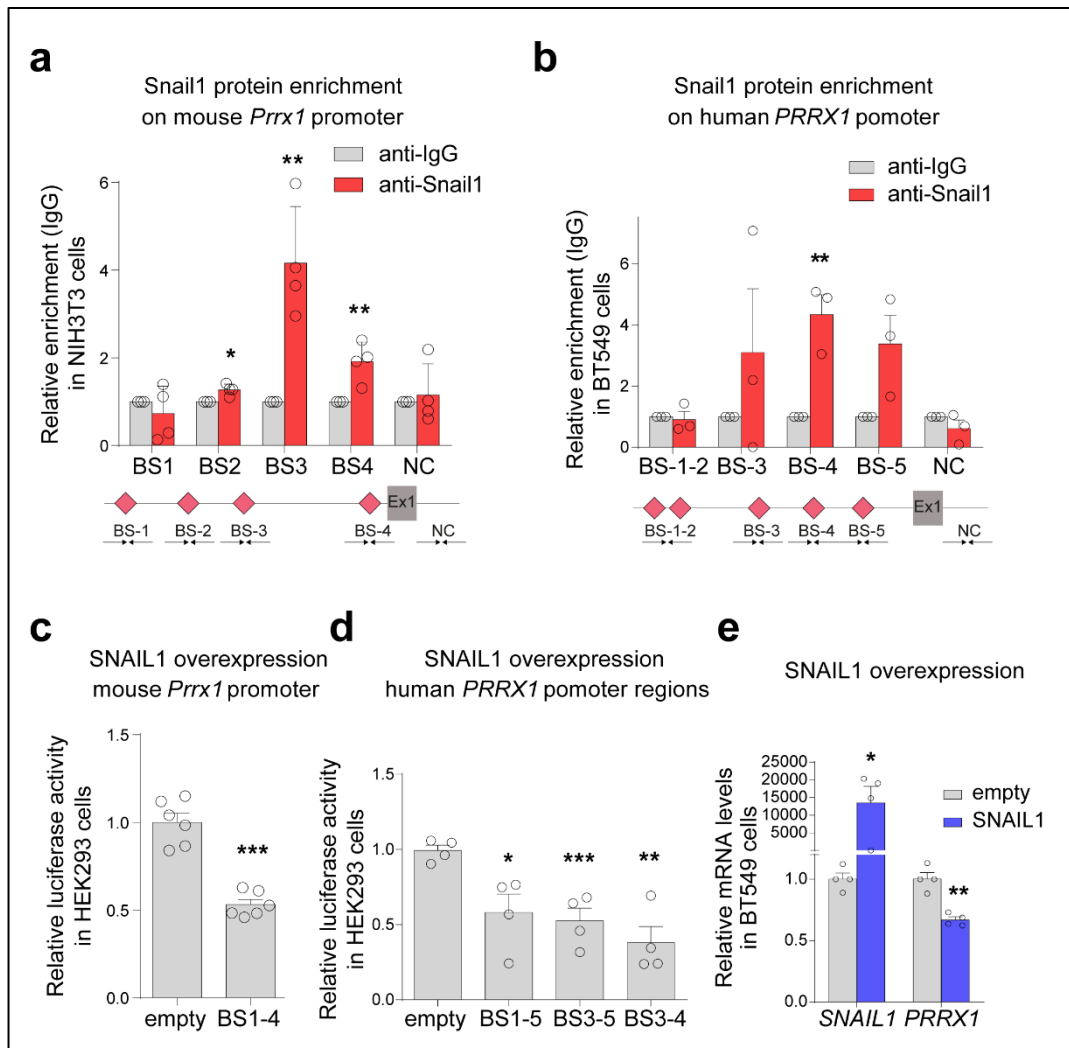


Figure 15. Snail1 directly represses *Prrx1* promoter in mouse and human cells. (a) Snail1 enrichment in the mouse *Prrx1* promoter shown by ChIP assay in NIH3T3 cells using a Snail1 specific antibody and IgG as control. A schematic map is shown, where red diamonds represent Snail1 potential binding sites (E-boxes; CANNTG) on *Prrx1* promoter. Designed primers are represented by arrows. Ex1: *Prrx1* exon 1. (b) SNAIL1 directly binds to the human *PRRX1* promoter, shown by ChIP assay in BT549 cells using a similar protocol. (c) Repression of mouse *Prrx1* promoter after SNAIL1 overexpression shown by luciferase assay in HEK293 cells. (d) Similar experiments using different regions of the human *PRRX1* promoter. (e) qPCR

assay showing downregulation of *PRRX1* transcription upon overexpression of SNAIL1 in BT549 cells.

BS: binding site and NC: negative control region. K: T/G, D: G/A/T, S: G/C and N: G/A/T/C. In every case, bars represent 2-6 independent experiments as biological replicates, represented by circles, and asterisks indicate significant p-value in t-test compared to the control in each case (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$).

To examine the impact of binding on promoter activity, we cloned both mouse and human *Prrx1* promoter regions upstream the Firefly Luciferase ORF in the pGL3 vector, and performed dual Luciferase Reporter assays after overexpressing SNAIL1 in human cells (HEK293). Mouse *Prrx1* promoter analysis shows a significant negative regulatory effect upon SNAIL overexpression (Figure 15c). For the human *PRRX1* promoter, we used the fragment containing all 5 binding sites and also fragments containing the sites where we found enrichment for binding. Likewise, the promoter activity is decreased upon SNAIL1 overexpression (Figure 15d). In agreement with this, *PRRX1* was downregulated after SNAIL1 transfection in BT549 cells, (Figure 15e). Altogether, these results indicate that Snail1 directly binds to, and represses *Prrx1* promoter activity both in mouse and human cells.

When searching for Snail1 E-boxes on *Prrx1* promoter, we also found several predicted *Prrx1* binding sites (TAATTGG, TAATTGC, AATTGC)(Xie et al., 2010) in both mouse and human sequences (Figure 16a-b). Thus, we next examined whether *Prrx1* could regulate its own transcription. We performed ChIP assays for endogenous *Prrx1* protein in NIH3T3 mouse fibroblast cells, and found two regions with enrichment for *Prrx1* occupancy in its own promoter (BS1: -4053, and BS3: -662, Figure 16a). In human BT549 cells, several regions of *PRRX1* promoter showed a tendency for enrichment of endogenous *PRRX1* protein occupancy as well, with BS1: -6875 being significantly enriched (Figure 16b). Luciferase assay for both mouse and human sequences show an activation of the *Prrx1* promoter upon *Prrx1* overexpression. Again, different sizes of the promoter containing different BSs were cloned to identify the regions of higher affinity for *PRRX1* binding (Figure 16c-d).

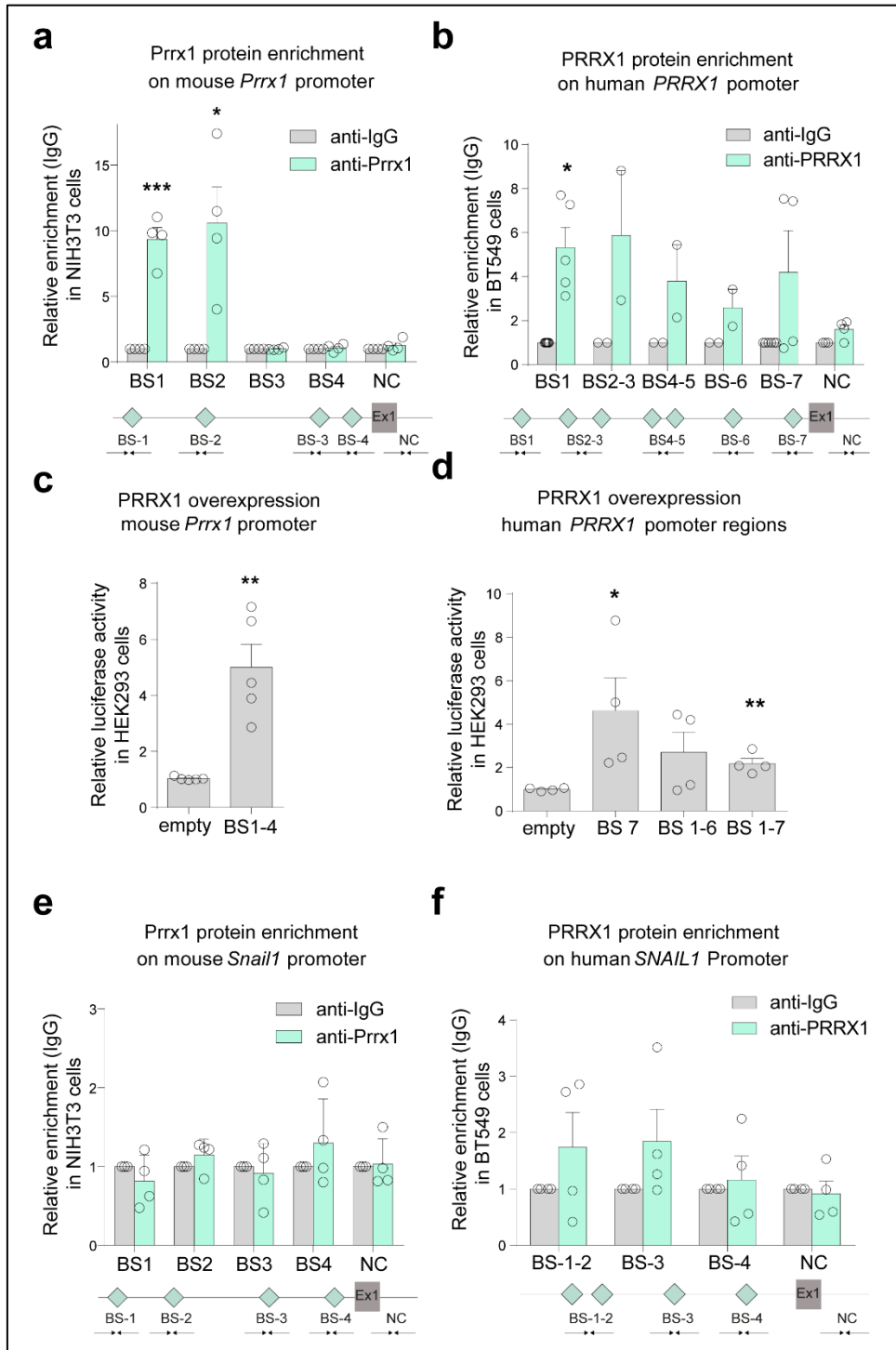


Figure 16. Prrx1 induces its own promoter in mouse and human cells, while it does not bind to *Snail1* promoter. (a) Prrx1 directly binds to its own promoter, as assessed by ChIP assays in NIH3T3 cells using Prrx1 specific antibody and IgG as control. A schematic map is shown, where cyan diamonds represent Prrx1 potential binding sites (TAATKDS) on its own promoter. The primers used are represented by arrows. Ex1: *Prrx1* exon 1. (b) PRRX1 directly binds to its own promoter shown by ChIP assay in BT549 cells with PRRX1 antibody and IgG as control. (c) Activation of mouse *Prrx1* promoter after Prrx1 overexpression in luciferase

assays in HEK293 cells. (d) Similar experiment with different regions of the human *PRRX1* promoter and PRRX1 protein. (e) Lack of enrichment for Prrx1 binding to mouse *Snail1* promoter in ChIP assays in NIH3T3 cells with a Prrx1 antibody and IgG as control. Cyan diamonds represent Prrx1 potential binding sites on *Snail1* promoter and primers are represented by arrows. Ex1: *Snail1* exon 1. (f) Lack of enrichment for PRRX1 binding to human *SNAIL1* promoter shown by ChIP assay in BT549 cells using PRRX1 specific antibody and IgG as control.

BS: binding site and NC: negative control region. K: T/G, D: G/A/T, S: G/C and N: G/A/T/C. In every case bars represent 2-6 independent experiments as biological replicates, represented by circles, and asterisks indicate significant p-value in t-test compared to the control in each case (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$).

Next, we wanted to know whether there is a reciprocal regulation between *Snail1* and *Prrx1*, and we examined if *Prrx1* was enriched in the *Snail1* promoter. However, although we found putative *Prrx1* binding sites within the *Snail1* proximal promoter region, ChIP assays in both mouse (Figure 16e) and human (Figure 16f) failed to show a significant *Prrx1* protein occupancy in those sites. Altogether, these results indicate that *Prrx1* is able to bind to its own promoter and enhance its own expression, but it does not bind to *Snail1* promoter, suggesting that *Prrx1* is not a direct *Snail1* transcriptional repressor. In fact, *Prrx1* has been described as a transcriptional activator (Grueneberg et al., 1992), which suggests a putative indirect regulation of *Snail1* by *Prrx1*.

4.3 *Prrx1* directly induces the transcription of *miR-15* family members

We had stable cell lines previously generated in the lab where human PRRX1 was ectopically expressed in MDA-MB-231 breast cancer cells, as they express *SNAIL1* and have PRRX1 at a very low level. Microarray (Affymetrix) analyses in these cells showed that *SNAIL1* was downregulated after PRRX1 overexpression, reinforcing the idea of a mutual negative regulation. Interestingly, we found several miRNAs that were upregulated in MDA231-PRRX1 cells, which could be potentially induced by PRRX1 (Figure 17a). Among those, we selected miR-424 and miR-503 (hereafter referred as

Mir-15-P1d and Mir-15-P2d (Fromm et al., 2015); Table IV) because they are conserved in different vertebrates, they have a relatively well described promoter, and most importantly, because their described functions as regulators of stemness, invasion, migration and cell proliferation, are compatible with those described for PRRX1 (Zhang et al., 2014, Drasin et al., 2015a, Li et al., 2014b). We validated the Microarray data by qPCR analyses, which confirmed their upregulation upon stable overexpression of PRRX1 in MDA231 cells (Figure 17b).

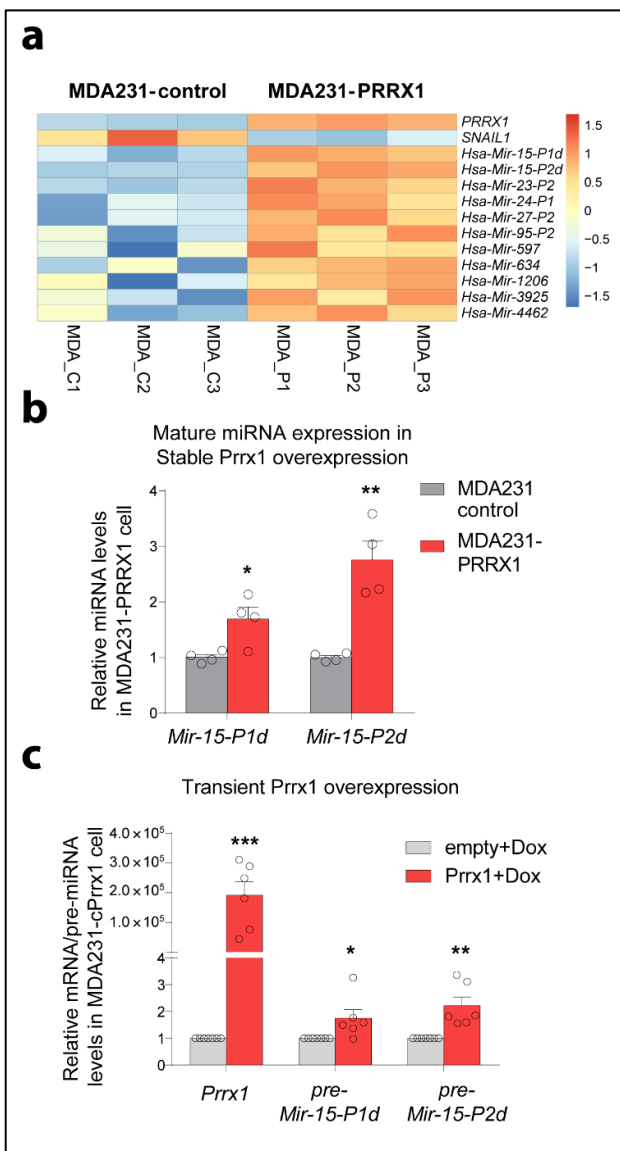


Figure 17. PRRX1 is upstream of miR-15 family members. (a) Heatmap showing Robust Multi-array Average (RMA)

normalized probe intensity values of Prrx1, Snail1 and selected miRNAs in MDA231 cells, with three control samples (MDA_C1 to C3) and three samples in which Prrx1 was overexpressed (MDA_P1 to P3). (b) Validation of *Mir-15-P1d* and *Mir-15-P2d* upregulation upon PRRX1 overexpression in MDA231-PRRX1 stable cell lines by Taq-Man qPCR using specific probes to detect the mature miRNAs. (c) *Prrx1*, *Mir-15-P1d_pre* and *Mir-15-P2d_pre* upregulation upon conditional (Dox-mediated) Prrx1 overexpression by qPCR in MDA231 cells.

cPrrx1: conditional Prrx1 overexpression; Dox: doxycycline. Bars represent 4-6 independent experiments as biological replicates, and asterisks indicate significant p-value in t-test. (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$).

To test whether the transcription of these miRNAs could be rapidly induced by PRRX1, we examined the expression of their precursor miRNAs (Pre-miRNAs) upon transient overexpression of mouse Prrx1 in MDA231 cells. To do so, we generated an inducible lentiviral system (Tet-ON) where nuclear yellow fluorescent protein (nYFP)-

P2A-Prrx1 can be temporally activated upon doxycycline (dox) treatment. 48h after dox induction, YFP positive cells were FACS sorted and subjected to qPCR, showing that Prrx1 itself and both Pre-miRNAs were significantly upregulated upon Prrx1 overexpression (Figure 17c), suggesting that they may be direct targets of Prrx1.

As miRNA molecules of the same family share the seed sequence and they can potentially target the same mRNAs, we examined whether the expression of other miR15 family members also correlated with that of PRRX1. Thus, we performed transient knock-down of human PRRX1 in BT549 cells with a specific short hairpin RNA (Ocana et al., 2012) (shRNA) vector containing nYFP. We then FACS sorted positive cells, and checked Pre-miRNAs expression for an early transcriptional response. We observed that *PRRX1* transcripts were significantly reduced, and all the members of the miR-15 family were significantly downregulated upon PRRX1 knock down (Figure 18a). We also confirmed this in HEK293 cells, where Pre-miRNAs of the miR-15 family were all significantly downregulated upon PRRX1 knock-down as well (Figure 18b).

To assess whether PRRX1 directly binds the *miR-15* family promoter to activate the expression of its members, we performed ChIP assays in BT459 cells using the human *Mir-15-P1/2d* promoter. We found enrichment for PRRX1 binding in at least one putative binding site in the promoter region, indicating that PRRX1 can directly bind to the human *Mir-15-P1/2d* promoter (Figure 18c). Next, the promoter region of the human *Mir-15-P1/2d* cluster was cloned upstream of the Firefly *luciferase* gene in the pGL3 vector. Luciferase assays showed that PRRX1 overexpression increases the activity of the human *Mir-15-P1/2d* promoter, and the deletion of PRRX1 potential binding sites within the promoter prevents this activation (Figure 18d). On the other hand, luciferase assays with *Mir-15-P1/2d* promoter and knock-down of PRRX1 in BT549 cells, shows a reduction in promoter activity, confirming that PRRX1 enhances *Mir-15-P1/2d* transcription (Figure 18e). We also performed ChIP and Luciferase assays in mouse NIH-3T3 cells and confirmed that this direct binding and activation of the *Mir-15-P1/2d* promoter by mouse Prrx1 protein is conserved between human and mouse (Figure 18 f-g). Furthermore, ChIP assays for Prrx1 in the promoter region of other members of the miR-15 family in human and mouse show that Prrx1 can directly bind and likely activate other members of the family as well (Figure 18h-j). Altogether, these experiments indicate that PRRX1 is able to directly bind to the promoter, and activate the transcription of miR15 family members.

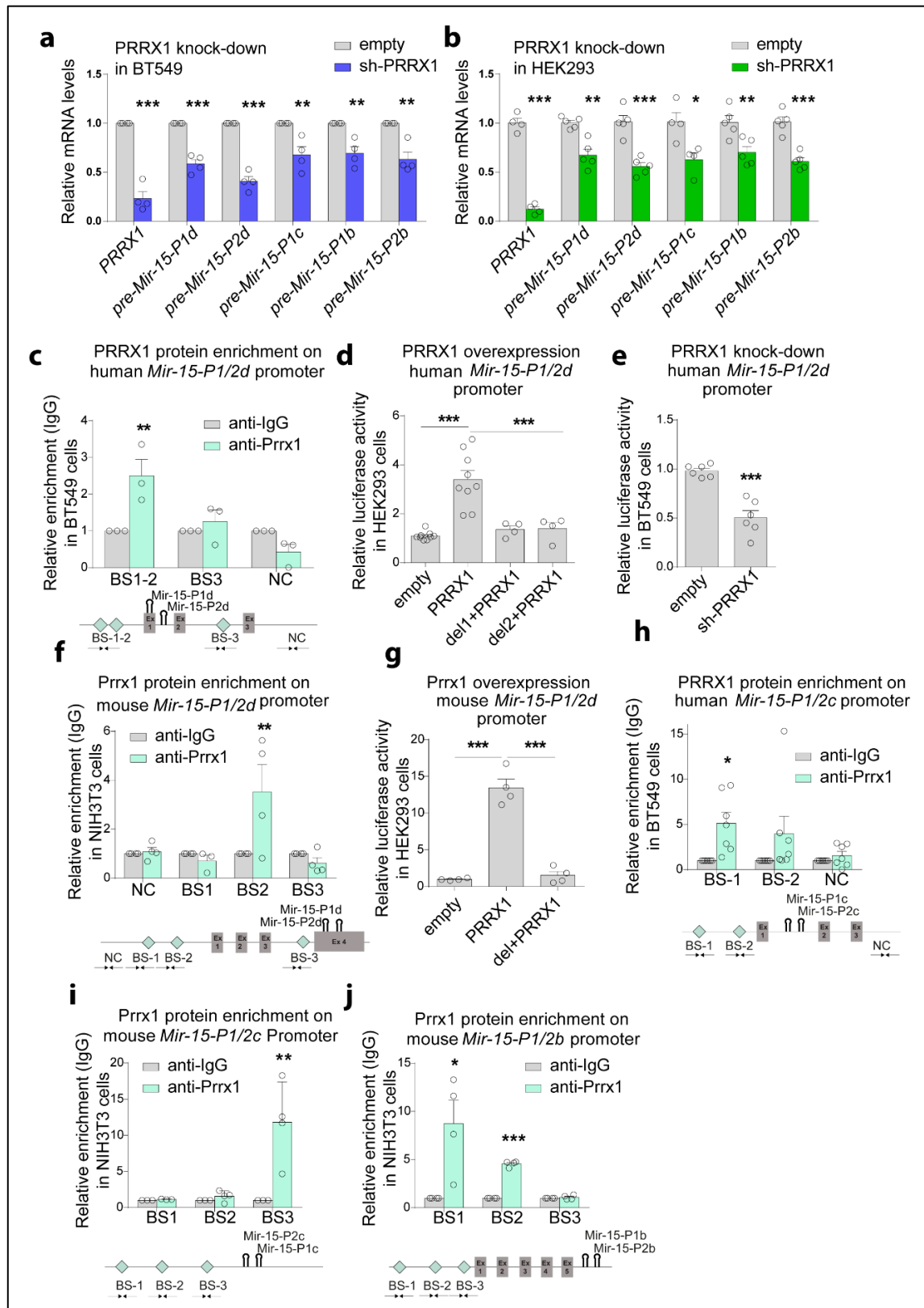


Figure 18. Prrx1 directly activates the transcription of the *miR-15* family. (a) qPCR assay showing downregulation of *PRRX1* and *premiRNAs* transcription upon transient knock-down of *PRRX1* in BT549 cells. *PRRX1* shRNA plus YFP transfected cells were sorted after 4 days. (b) Similar assays in HEK293 cells. (c) *PRRX1* directly binds to the human *Mir-15-P1/2d*

promoter as shown by ChIP assay in BT549 cells using PRRX1 specific antibody and IgG as control. Cyan diamonds in the schematic map represent PRRX1 potential motifs, TAATKDS, on *Mir-15-P1/2d* promoter. Ex1-3 represent exons of the host long non-coding RNA *MIR503HG*. Arrows represent primer sets used for ChIP detection. **(d)** Activation of human *Mir-15-P1/2d* promoter by PRRX1 overexpression shown by dual luciferase assay in HEK293 cells. This activation is abolished upon deletion of the PRRX1 binding sites in *Mir-15-P1/2d* promoter (del1/2+PRRX1). **(e)** Repression of human *Mir-15-P1/2d* promoter upon knock-down of PRRX1 by specific shRNA shown by dual luciferase assay in BT549 cells. **(f)** PRRX1 directly binds to the mouse *Mir-15-P1/2d* promoter shown by ChIP assay in NIH3T3 cells using Prrx1 specific antibody and IgG as control. **(g)** Activation of mouse *Mir-15-P1/2d* promoter by Prrx1 overexpression shown by dual luciferase assay in HEK293 cells. This activation is abolished upon deletion of Prrx1 binding site in *Mir-15-P1/2d* promoter. **(h-j)** Prrx1 directly binds to the human *Mir-15-P1/2c*, mouse *Mir-15-P1/2c* and *Mir-15-P1/2b* promoters shown by ChIP assay in BT549 and NIH3T3 cells, respectively, using Prrx1 specific antibody and IgG as control.

BS: binding site and NC: negative control region. K: T/G, D: G/A/T and S: G/C. Bars represent 3-6 independent experiments as biological replicates, and asterisks indicate significant p-value in t-test or one-way ANOVA with Bonferroni's multiple comparison test compared to the control and other samples in each test, where appropriate (* p < 0.05, ** p < 0.01 and *** p < 0.001).

Table IV, miR-15 family nomenclature and accession numbers	
miR-15 family	MIPF0000006
<i>Has-Mir-15-P1d (miR-424)</i>	MI0001446
<i>Has-Mir-15-P2d (miR-503)</i>	MI0003188
<i>Has-Mir-15-P1b (miR-15b)</i>	MI0000438
<i>Has-Mir-15-P2b (miR-16)</i>	MI0000115
<i>Has-Mir-15-P2c (miR-195)</i>	MI0000489
<i>Has-Mir-15-P1c (miR-497)</i>	MI0003138

4.4 miR-15 and Prrx1 share expression sites in embryonic development and correlate with survival in cancer patients

Once the direct induction of the miR-15 family by Prrx1 was confirmed in cultured cells, we analyzed the expression of these miRNAs during embryonic development and also in cancer. At E8.5, where the complementary expression of Snail and Prrx1 in developing mouse embryo is evident in MNC and LPM cells (Figures 13a-b), we studied the expression of the miRNAs by ISH using specific DIG-labeled LNA (locked nucleoid acid) probes to detect their mature versions. Expression can be detected in similar territories to those of *Prrx1*. See examples in the cranial region in Figure 19a, in agreement with Prrx1 inducing the expression of these miRNAs.

As shown previously, Snail1 high expression correlate with poor prognosis (Ye et al., 2015, Ocana et al., 2012), while PRRX1 high expression correlates with good prognosis among breast cancer patients (Ocana et al., 2012). We have confirmed this notion analyzing breast cancer patients' data from another dataset (Gyorffy et al., 2010) containing all different conditions and subtypes (Total) (Figure 19b). We checked whether the expression of *miR-15* family members have similar profile to that of *PRRX1* in terms of patients' overall survival (OS). From the same dataset, we analyzed the expression of *miR-15* family members, and found similar pattern to that of *PRRX1* i.e. high expression of *miR-15* family members correlated with higher OS (Figure 19b). Interestingly, when these analyses were done for patients with lymph node positive status, the correlations observed for *PRRX1* and *miR-15* family members was even more significant (Figure 19b). In line with previous finding that *PRRX1* high expression correlated with better OS, patients have significant better survival when *PRRX1/miR-15* family expression is high. See discussion for a reasoned explanation of the positive correlation between the association of high Prrx1 expression and both invasion and patients' survival. Altogether, these results indicate that *miR-15* family expression is positively correlated with *Prrx1* in development, and also when both are expressed at high levels in cancer patients, they are associated with better prognosis.

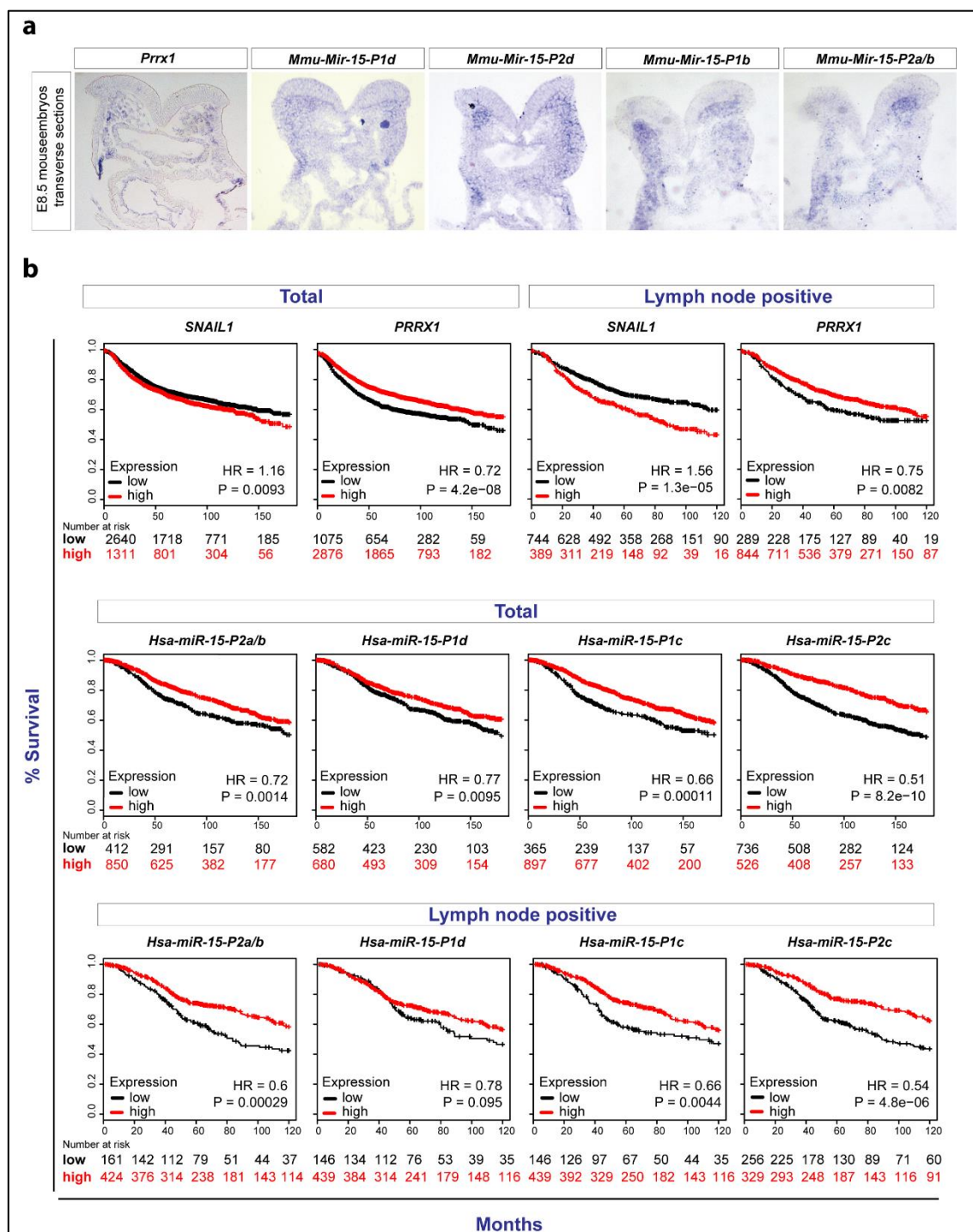


Figure 19. Expression of miR-15 family members correlates with *Prrx1* in mouse embryo and survival of cancer patients. (a) Transverse sections of the cranial region of E8.5 embryos showing the expression of *Prrx1*, *Mir-15-P1d*, *Mir-15-P2d*, *Mir-15-P1b* and *Mir-15-P2a/b* in similar regions. (b) Kaplan-Meier overall survival (OS) plots from breast cancer patients showing that high expression of *SNAIL1* correlates with low survival, while *PRRX1* high expression correlates with a better survival. Similarly, the expression data of *miR-15* family members follow a similar trend as that of *PRRX1*. Hazard ratio (HR) and logarithmic ranked P Value (longrank P) were analyzed to infer the significance of the differences.

4.5 Prrx1 attenuates *Snail1* expression through miR-15 family

We validated our microarray data (Figure 17a) by qPCR, and confirmed that in MDA231 cells stably overexpressing PRRX1, *SNAIL1* is significantly downregulated (Figure 20a). To assess whether this was an early response, we examined *SNAIL1* expression in an inducible Prrx1 system. We analyzed MDA231-nYFP-P2A-Prrx1 cells after 48hs of induction with Doxycycline and found that *SNAIL1* mRNA was downregulated upon PRRX1 induction, confirming the microarray data also in a short-term experiment (Figure 20b).

The observation of a negative correlation of *SNAIL1* after constitutive and short-term induction of PRRX1 expression together with that of PRRX1 directly inducing the transcription of *miR-15* family members which are also predicted to target *SNAIL1* (Figure 20c), made us to test the possibility that Prrx1 could be attenuating *SNAIL1* expression through the activation of miR-15 family members. Importantly, the miRNA responsive element (MRE) in *Snail1* 3'UTR is conserved in all vertebrates (Figure 20d). Thus, we examined whether *SNAIL1* could be directly targeted by miR-15 family miRNAs. We first overexpressed *Mir-15-P1/2d* in MDA231 cells, and found a significant downregulation of *SNAIL1*. As a control, we overexpressed a seed-mutated version of the miRNAs (*Mir-15-P1/2d-mut*), unable to bind to *SNAIL1* 3'UTR (Figure 20e). Conversely, we used specific miRNA inhibitors/LNAs to knock-down *Mir-15-P1/2d* in BT549 cells, which express high levels of PRRX1 and miR-15, and that resulted in upregulation of *SNAIL1* (Figure 20f). To confirm that this downregulation was through direct binding of the miRNAs to the 3'UTR, we performed Luciferase assays and observed that overexpression of *Mir-15-P1/2d* repressed human *SNAIL1* 3'UTR activity. This repression was abolished when we used either the seed-mutant control of the miRNAs or a MRE-mutated version of the *SNAIL1* 3'UTR (Figure 20g). Moreover, we checked other members of the miR-15 family and found that *Mir-15-P1/2b* behaves similarly (Figure 20h). This regulation is conserved in mouse, chicken and zebrafish, as overexpression of *miR-15* family members can repress *Snail1* 3'UTR in those species (Figure 20i-k), indicating that *Snail1* is a bona fide target of the miR-15 family in vertebrates.

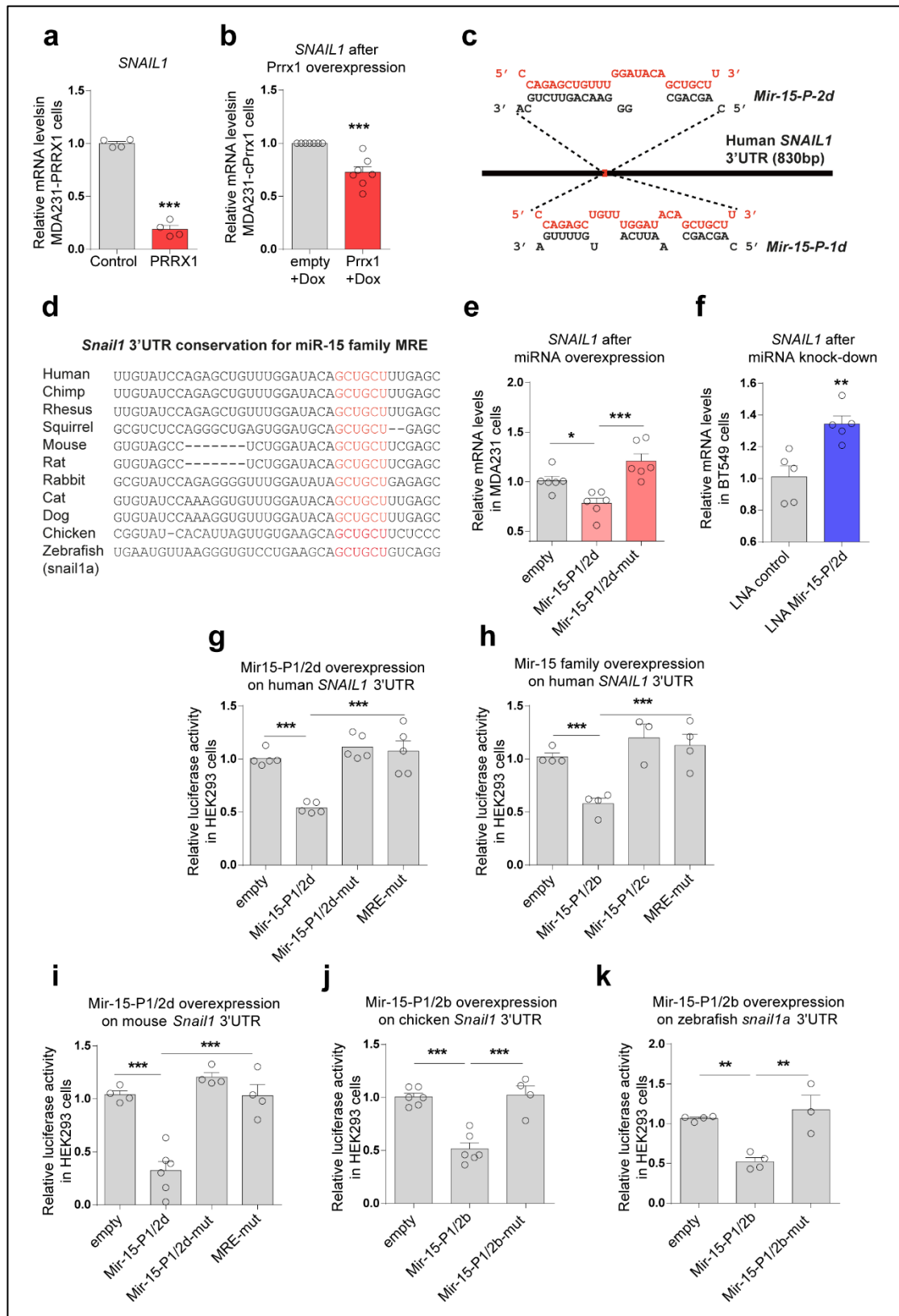


Figure 20. miR-15 family directly binds to *SNAIL1* 3'UTR and attenuates its expression. (a) qPCR assays upon stable overexpression of PRRX1 confirms the observed downregulation of *SNAIL1* in microarrays. (b) qPCR assay showing the downregulation of *SNAIL1* upon transient overexpression of Prrx1 in MDA231 cells infected with inducible nYFP-P2A-Prrx1.

RESULTS

Cells were sorted 48 hours after induction with doxycycline. **(c)** Schematic representation of *SNAIL1* 3'UTR, predicted to be a target of *Mir-15-P1/2d*, by RNAhybrid (<https://bibiserv2.cebitec.uni-bielefeld.de/rnahybrid>). **(d)** Conservation of miR-15 family binding sites in *SNAIL1* 3' UTR sequences in vertebrates. **(e)** qPCR assay showing downregulation of *SNAIL1* upon transient overexpression of *Mir-15-P1/2d* in MDA231 cells. This downregulation does not occur after overexpression of miRNAs versions with a mutant seed sequence (*Mir-15-P1/2d-mut*). **(f)** qPCR assay showing upregulation of *SNAIL1* upon transient knock-down of *Mir-15-P1/2d* in BT549 cells using miRNA inhibitors. **(g-k)** Luciferase assays in HEK293 cells. **(g)** Repression of human *SNAIL1* 3' UTR after *Mir-15-P1/2d* overexpression. This repression is not observed when mutant versions of the seed sequence (*Mir-15-P1/2d-mut*) or of the MRE in *SNAIL1* 3-UTR (MRE-mut) were used. **(h)** Repression of human *SNAIL1* 3'UTR by other *miR-15* family members. This repression is abolished upon mutation of *MRE* in *SNAIL1* 3-UTR. **(i-k)** Similar experiments to those shown in (g), but using either the mouse or the chicken *Snail1* or the zebrafish *snail1a* 3'UTRs, respectively. This repression is abolished upon mutation in the seed sequence of miRNAs.

MRE: miRNA responsive element. Bars represent 4-6 independent experiments as biological replicates, and asterisks indicate significant p-value in t-test or one-way ANOVA with Bonferroni's multiple comparison test compared to the control and other samples in each test, where appropriate (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$).

To check *SNAIL1* protein levels upon mouse *Prrx1* overexpression, we took advantage of the conditional MDA231-nYFP-P2A-*Prrx1* system, where we could see YFP (*Prrx1*) positive and negative nuclei and correlate that with *SNAIL1* expression. After *Prrx1* induction, the overall expression of *SNAIL1* decreases dramatically, and *SNAIL1* positive nuclei are mostly found in YFP negative cells ([Figure 21a](#)), which confirms *SNAIL1* mRNA downregulation mainly in those cells where *Prrx1* is overexpressed ([Figure 20b](#)). We quantified the intensity of the *SNAIL1* signal in YFP⁻ (*Prrx1*⁻) cells and in those overexpressing *Prrx1* and found a significant downregulation in the latter ([Figure 21b](#)).

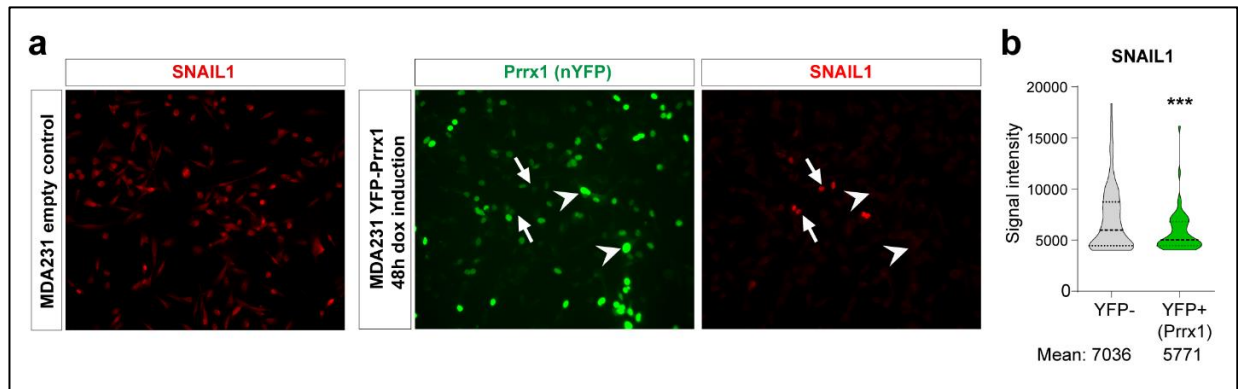


Figure 21. Repression of SNAIL1 protein upon mouse Prrx1 overexpression in breast cancer cells. (a) Complementary expression of SNAIL1 and Prrx1 in MDA231 breast cancer cells infected with conditional nYFP-P2A-Prrx1. SNAIL1 (red) and Prrx1 (green) positive nuclei. Cells expressing high levels of Prrx1 are devoid of SNAIL1 expression (examples highlighted with arrowheads). Arrows indicate examples of cells that failed to activate high Prrx1 expression and maintain SNAIL1 expression (b) Quantification of relative intensities. Intensity analyses for SNAIL1 (red signal) shows a significant reduction in nYFP (Prrx1) positive cells.

To test whether this regulation operates *in vivo*, we knocked down *miR-15* family in zebrafish embryos by injecting a sponge RNA (Ebert et al., 2007) containing complementary sequences to those of *miR-15* family members. Sponge RNAs quench and/or degrade miRNAs with complementary sequences, and are good tools in particular when aiming at targeting several members of a family, as they share the seed sequence (and likely additional nucleotides) (Figure 22a). Zebrafish embryos were injected at the 1-2 cell stage and then collected at around the 20-somite stage and subjected to ISH for *snail1a/b*. In one third (30/96) of the injected embryos, *snail1a/b* expression was increased, in some cases inducing severe morphological malformations compatible with the induced ectopic position of cells (Figure 22b, asterisk). Transverse sections confirm that upon Sponge injection, there is a drastic increase in *snail1* expression in the territories of endogenous *miR-15* expression (migratory neural crest/ head mesenchyme) (Figure 22c). To quantify these results, we collected pools of embryos and performed qPCR, and confirmed that both *snail1a* and *snail1b* transcripts are more abundantly expressed in zebrafish embryos injected with *miR-15* family sponge (Figure 22d). These findings are compatible with the roles of *miR-15* family attenuating *snail1* expression in zebrafish, confirming the results previously observed in cultured cells (Figures 20-21).

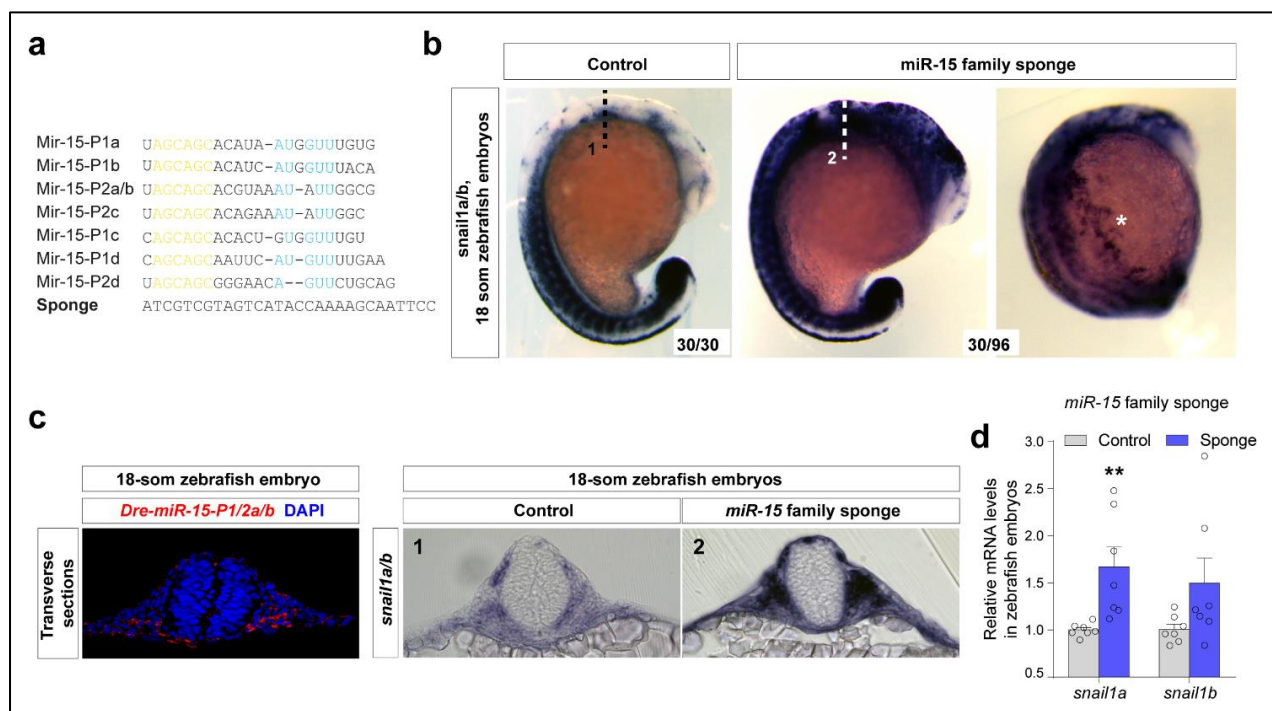


Figure 22. *snail1* attenuation through *Prrx1*-induced miR-15 family in zebrafish. (a) Sequence alignment for members of the miR-15 family, and the complementary sequence of the Sponge used. The seed sequence is highlighted in orange and blue nucleotides indicate additional conserved residues within family members. (b) 20-somite control or *miR-15* family sponge-injected zebrafish embryos showing *snail1a/b* expression by in situ hybridization. (c) Fluorescent miRNA *in situ* hybridization using specific LNA probes to detect mature miRNAs in transverse sections of 20 somite zebrafish embryo (Left), showing the expression pattern of *Dre-miR-15-P1/2a/b* in neural crest derived cell populations. Transverse sections taken from the embryos shown in (b) at the levels indicated by the dashed lines (Right). (d) qPCR analysis shows upregulation of *snail1a/b* expression in the sponge-injected embryos compared to controls.

We also assessed whether this regulation was conserved in the mouse embryo, and we performed western blot analysis from embryos bearing a LacZ knock-in in the *Prrx1* locus (*Prrx1^{tm1Jfm}*) (Lu et al., 1999) (Figure 23a). Quantification of signal intensity indicate that there is a significant increase of Snail1 protein levels in mutant embryos (Figure 23b). We performed Snail1 IF in cryo-sections of E11.5 WT and mutant mouse embryos (Figure 23c-d) and we found an increase in expression and an expansion of Snail1 territories in mutant embryos, as expected considering the mutual repression that we describe here. As an example, here we show a transverse section at the head

level showing Snail1 expansion in the nasal pit (Figure 23c) and in the medio-lateral nasal process (Figure 23d), evidenced by extension of positive cells and confirmed after quantification of signal intensity (Figure 23c-d). Altogether, these findings are compatible with our *in vitro* analyses and indicate that Prrx1 indirectly attenuates the expression of Snail1, through the induction of miR-15 family, and that this mechanism is conserved in different vertebrates.

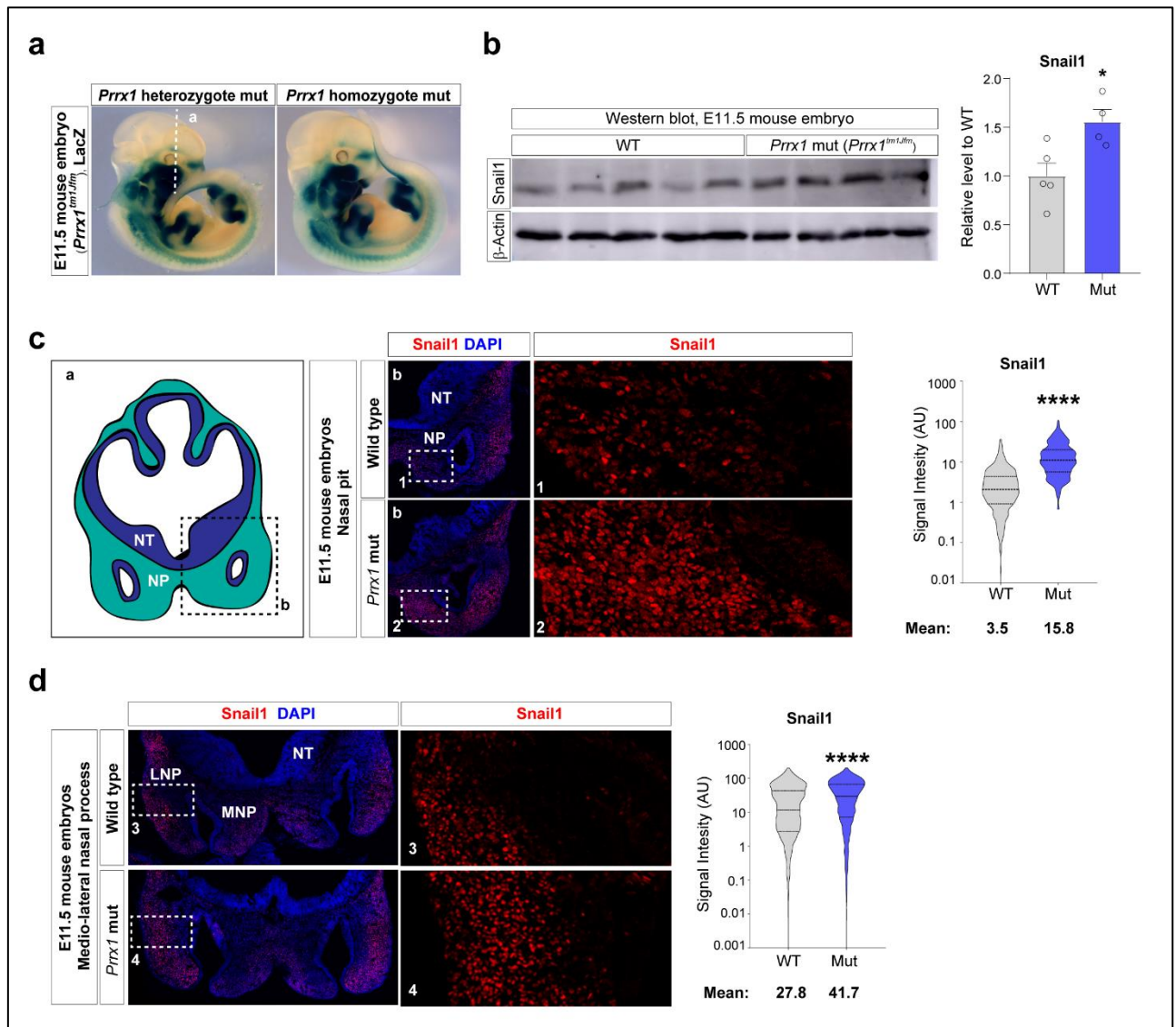


Figure 23. Repression of Snail1 through Prrx1-induced miR-15 family in the mouse. (a) LacZ staining of E11.5 mutant mouse embryos in hetero or homozygous embryos, (*129S-Prrx1^{tm1Jfm}*). (b) Western blot for Snail1 and β-actin from E11.5 control and Prrx1 mutant embryos. (c) Quantification of intensity of Snail1 signal in western blot shows a significant increase of expression in mutant embryos compared to WT. (d, e) Snail1 IFs in the nasal pit (d) or nasal process (e) regions of E11.5 WT and mutant embryos show increased level as well

as expansion of Snail1 expression domains. Signal intensity for Snail1 IF (right panels in c and d) is measured calculating the average signal in the nuclei of all cells in the selected regions.

WT: wild type; Mut: mutant; NT: neural tube, NP: nasal pit; MNP: medial nasal process, LNP: lateral nasal process; AU: arbitrary units asterisks indicate significant p-value in t-test (* $p < 0.05$ and *** $p < 0.001$). Quantifications are performed for one section of WT or mutant embryos, and the increase and expansion is observed in 3 different E11.5 WT and 2 mutant embryos.

4.6 Snail1 and Prrx1 are sequentially expressed during developmental EMT program

Snail1 seems to be the first EMT-TF to be expressed in regions of the embryo that will undergo an EMT process, including the neural crest and the primitive streak (Sefton et al., 1998). *Snail* genes (*Snail1* in the mouse and *Snail2* in the chick) are expressed in the PNC before delamination from the neural tube and continue to be expressed in early migratory cells but their expression decays along the migratory routes (Sefton et al., 1998, Del Barrio and Nieto, 2004). In contrast, Prrx1 is expressed in migratory crest subpopulations (MNC; Ocaña et al., 2012 and this work). The expression of these two EMT-TFs follows a temporal order in the migration of the neural crest, and thus, likely reflects a hierarchy of activation during the EMT process that drives neural crest delamination and migration. To demonstrate this, we used a mouse model generated in the lab in which we could visualize this sequential expression (unpublished). The *Snail1* (DES)-GFP reporter mouse is a transgenic containing a *downstream enhancer* of *Snail1*, that drives the expression of GFP to its endogenous territories (Figure 24a), recapitulating the vast majority of *Snail1* expression in the developing embryo. As expected, the half-life of the GFP protein is longer than that of Snail1 in the mouse embryo (Figure 24b), and thus, this model provides a short-time lineage tracing system. In transverse sections of E8.5 embryos, we observed that premigratory and early MNC cells expressed Snail1 and GFP, while late MNC cells are negative for Snail1 protein but positive for both GFP and Prrx1 (Figure 24b). This observation is compatible with the idea that Snail1 and Prrx1 are expressed in a sequential manner, as Prrx1 positive cells seem to be descendants of Snail positive cells (double labeled cells in Fig. 23b). Altogether, these results suggest that during embryonic development, cells undergo

EMT by activating Snail1 first, and then attenuation of Snail1 is indirectly mediated by Prrx1 though the activation of the miR15 family, all being part of a GRN in which miR15 family may coordinate the transition from Snail1- to Prrx1-mediated EMT programs.

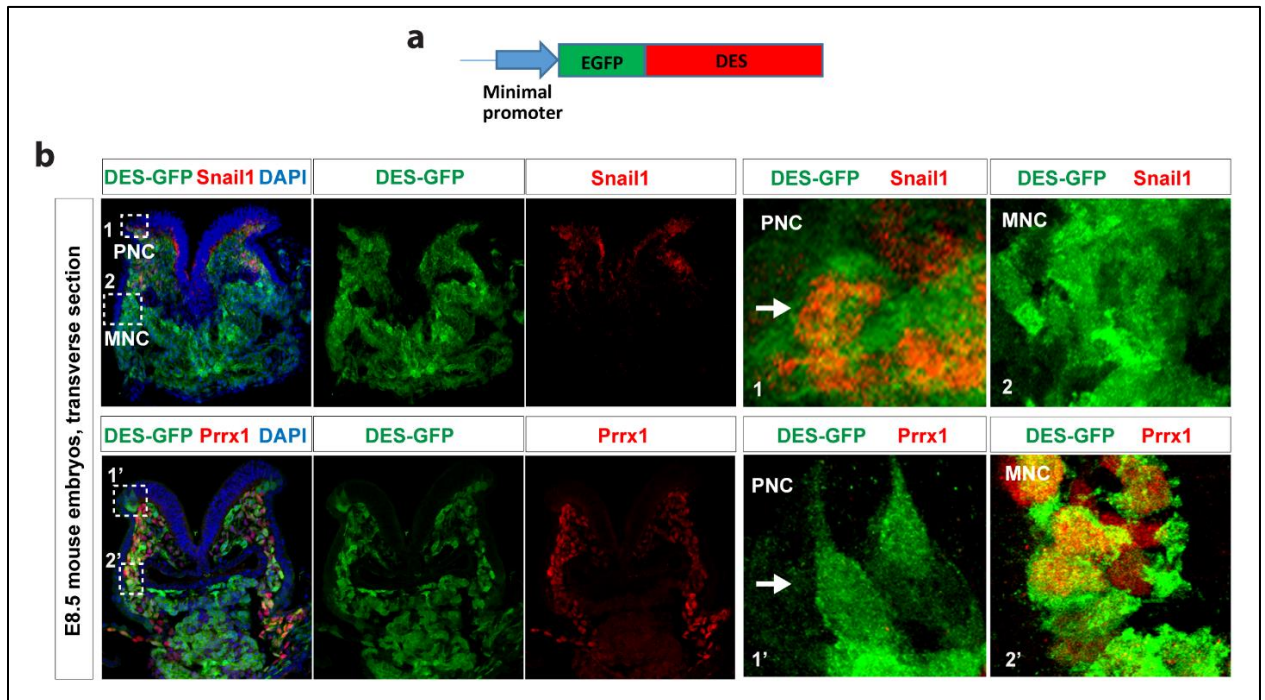


Figure 24. Snail1 and Prrx1 are sequentially expressed during EMT. (a) Schematic representation of the construct used to generate the transgenic mouse model. The downstream enhancer of Snail1 (DES) was cloned downstream of a minimal promoter and EGFP. (b) Transverse sections of E8.5 DES-GFP reporter mouse embryos showing IFs for GFP and Snail1 (upper panel) or Prrx1 (lower panel). Insets show higher magnification pictures for GFP and Snail1 (1 and 2), and for GFP and Prrx1 (1' and 2').

4.7 PRRX1 and miR-15 in the progression towards a full EMT

To better characterize the temporal regulation of Snail1 and Prrx1 expression and their contribution to EMT dynamics, we used MDCK cells as a substrate to induce and follow EMT with time. MDCK cells are derived from dog epithelial kidney cells, and have been extensively used for EMT studies (Thiery et al., 2009). Unpublished data from the lab indicate that different versions of MDCK cells have different competence to undergo EMT in response to TGF β . As such, MDCK-II and MDCK-NBL cells undergo partial and full EMT, respectively (Figure 25a, b, e, f). Interestingly, we have observed a different

dynamics of *SNAIL1* and *PRRX1* expression upon TGF β -mediated EMT induction in these two cell lines (Figure 25c, d). *SNAIL1* is induced very rapidly in both, and *PRRX1* is induced later, in line with the observation in mouse embryos. However, *SNAIL1* is more potently induced in MDCK-II cells, while *PRRX1* induction is much more potent in MDCK-NBL cells (Figure 25c, d). High levels of *PRRX1* expression are concomitant with a complete loss of epithelial identity in MDCK-NBL cells, compatible with these cells undergoing a full EMT (Figure 25b, f). In MDCK-II cells, even after a long treatment, epithelial markers such as the tight-junction protein-1 (TJP1) are still expressed concomitant with mesenchymal markers such as fibronectin (FN1), and cells do not undergo full mesenchymalization and stay in contact with each other (Figure 25e). This state is recognized as a partial EMT.

We then checked whether the expression of *miR-15* family members follow that of *PRRX1* in these cell lines. We checked the expression of mature miRNAs using TaqMan qPCR assay with specific probes for dog *miR-15* family members. Interestingly, miRs are not significantly upregulated in MDCK-II cells even after long treatments with TGF β , likely due to the decay in *PRRX1* expression (compare Figure 25c with g). By contrast, at least three members of the *miR-15* family are significantly upregulated after long treatments in MDCK-NBL cells, consistent with the presence of high levels of *PRRX1* and lower levels of *SNAIL1* (compare Figure 25d with h). We wondered whether this weak upregulation of *SNAIL1* in MDCK-NBL cells was due to the high *PRRX1* levels which in turn could activate the expression of *miR-15* family members and indirectly attenuate *SNAIL1* transcription (Figure 25c, d). Thus, we knocked-down *PRRX1* in MDCK-NBL cells concomitant with TGF β administration and we observed downregulation for all mature *miR-15* family members (Figure 25i), and *SNAIL1* upregulation. Under these circumstances, cells fail to undergo EMT (Figure 25j).

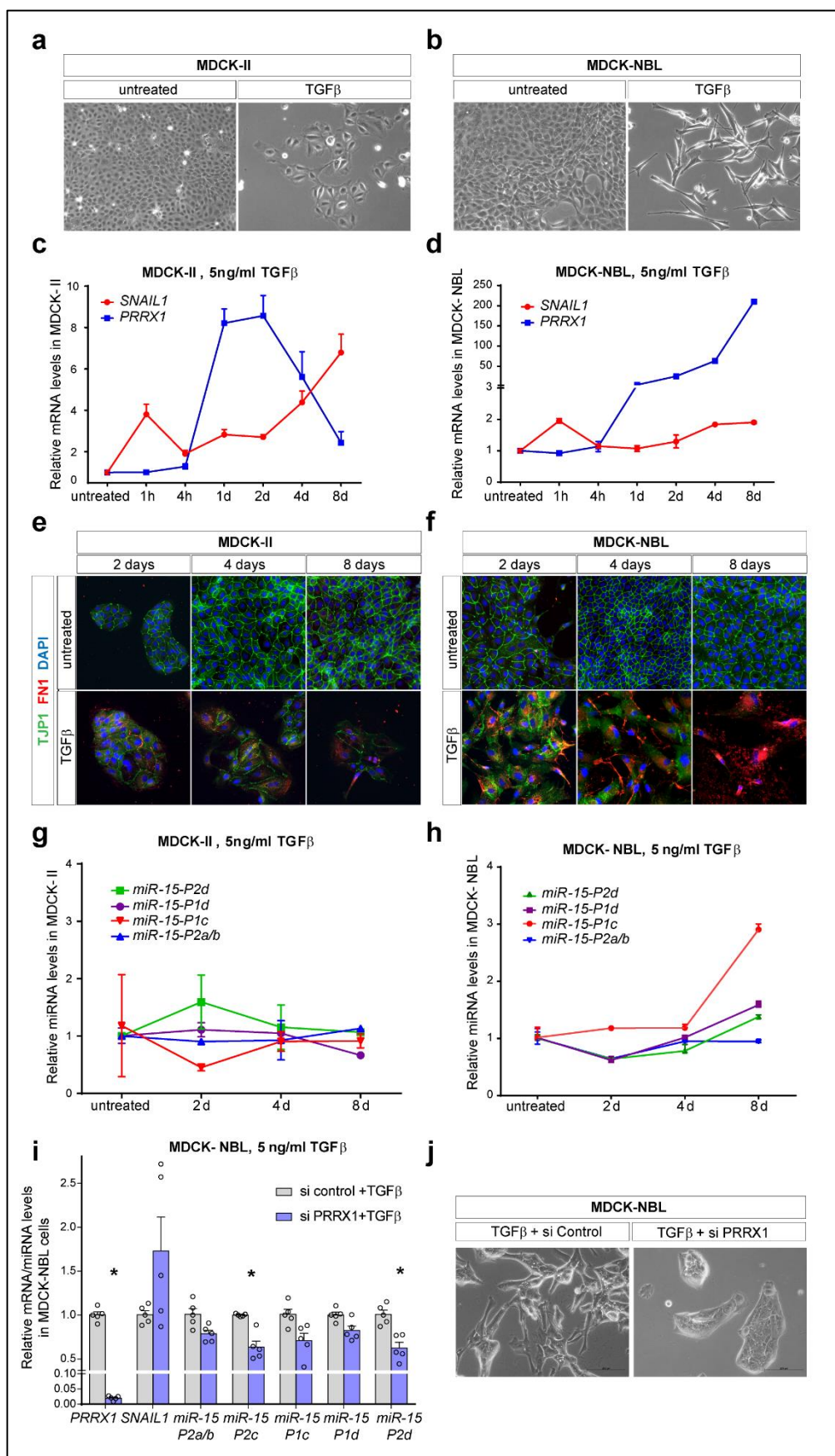


Figure 25. Activation of miR-15 family members in full-EMT upon PRRX1 potent

upregulation but not partial-EMT in MDCK cells. (a, b) Two derivatives of a parental MDCK cell line, MDCK-II and MDCK-NBL, behave differentially after TGF β administration, undergoing partial and full EMT, respectively. Cells were treated with 5ng/ml for the indicated time periods. (c) MDCK-II cells respond to TGF β with a rapid activation of *SNAIL1* at 1hour (1h, first peak) and a continuous increase in levels after an early decay. *PRRX1* was activated later but its expression dramatically decreased after two days of treatment. (d) Analogous experiments carried out in MDCK-NBL cells show a similarly rapid activation of *SNAIL1* (first peak) and a weak upregulation at later time point. *PRRX1* expression was activated later, with its expression increasing with time to very high levels. (e, f) IF for fibronectin (FN1) as readout of the acquisition of mesenchymal character and tight junction protein 1 (TJP1) as an epithelial marker. MDCK-II cells do not lose the epithelial marker even at 8 days after TGF β treatment. However, in MDCK-NBL cells, the epithelial marker is completely lost concomitant with a rapid gain of FN1. (g, h) Expression of mature *miR-15* family members in the two cell lines. miRs were not upregulated in MDCK-II cells, while in MDCK-NBL cells at least three members were upregulated after 8 days of treatments. (i) Expression of *PRRX1*, *SNAIL1* and mature *miR-15* family members in MDCK-NBL cells treated with TGF β in the absence or the presence of *PRRX1* siRNA. Results show that upon *PRRX1* downregulation, the expression *miR-15* family members are not activated and *SNAIL1* is upregulated. (j) MDCK-NBL cells after treatment with TGF β in the absence or the presence of siPRRX1. When PRRX1 is not activated, cells fail to undergo EMT. All the data in the figure are normalized with respect to the control value at each time point.

Altogether, these results suggest that for a full EMT induction, PRRX1 needs to reach high levels of expression, which in turn results in SNAIL1 attenuation through the activation of miR-15 expression. Interestingly, this also concurs with the late activation of the other potent epithelial repressor Zeb1 (unpublished data from the lab).

4.8 Snail1 but not Prrx1 is activated during partial EMT kidney fibrosis

Previous data from the lab showed that a partial EMT in epithelial kidney cells leads to renal fibrosis and organ failure in the unilateral ureteral obstruction (UUO) model (Grande et al., 2015). This concurred with the dedifferentiation of renal epithelial cells that could not function leading to renal insufficiency. These damaged cells secreted chemokines and cytokines that promoted fibrogenesis and inflammation. Importantly, those cells did not engage into the invasion program and stayed integrated in the damaged tubules. As they did not undergo a full EMT, we wondered whether this partial EMT was accompanied by the activation of Snail and the failure to activate Prrx1.

We performed UUO in WT mice to induce renal fibrosis and found that similarly to *Snail1*, *Prrx1* mRNA was also upregulated in a time-dependent manner in whole fibrotic kidney lysates when compared to control conditions (Figure 26a). However, when we looked for Prrx1 protein expression in histological sections, we found that, in contrast to Snail1 protein (Grande et al., 2015), which is expressed by both epithelial and stromal cells, Prrx1 was exclusively expressed in the expanded stromal cell population of fibrotic kidneys (Figure 26b). This explains the increase in *Prrx1* mRNA levels in the fibrotic kidneys. Furthermore, when we checked the expression of mature miR-15 family members by Taq-Man qPCR, some members were upregulated, consistent with the increase in Prrx1 mRNA levels (Figure 26c). Altogether, our results *in vivo* in a model of renal fibrosis are compatible with the failure to undergo a full EMT, as we have observed in MDCK cells when PRRX1 was not readily activated (Figure 25c). This raises the question of whether a full-EMT state requires Prrx1 activation. Out of the frame of this work, we plan to generate a model to specifically address this question. As such, we will generate a transgenic mouse line with ectopic expression of Prrx1 specifically in renal epithelial cells. We will knock-in an inducible *Prrx1* coding sequence which expression will be driven by the renal epithelial cell specific promoter (ksp) and will induce fibrosis by UUO as we have previously described (Grande et al., 2015). This is the perfect model to test whether Prrx1 upregulation in a Snail positive background can trigger the transition of epithelial cells from partial to full EMT.

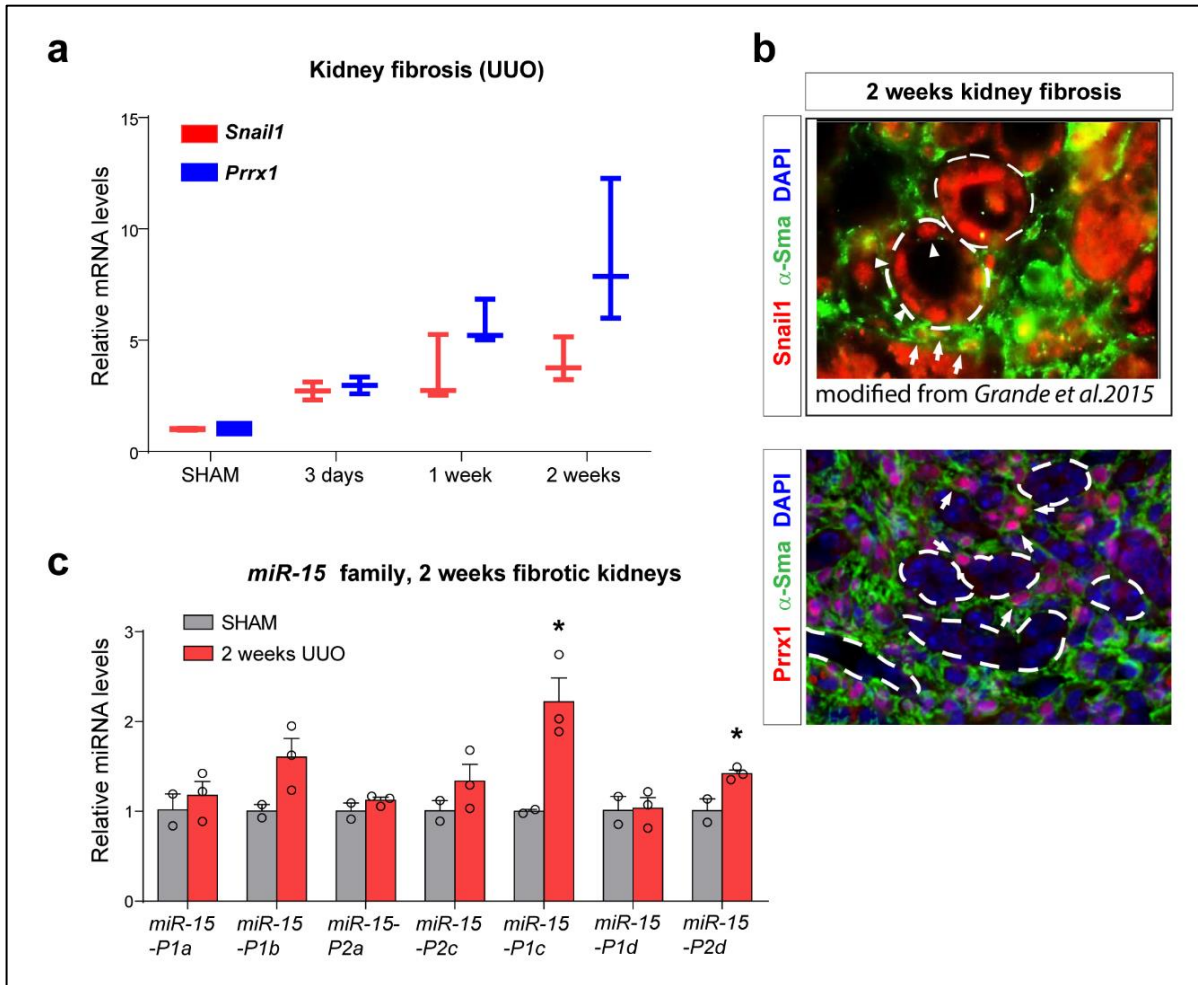


Figure 26. Snail1 but not Prrx1 is activated in renal epithelial cells during the partial EMT associated with kidney fibrosis. (a) Dynamics of *Prrx1* and *Snail1* expression in obstructed mouse kidneys. SHAM kidneys were used as control of those subjected to UUO. (b) (Top) IF for Snail1 and alpha-Smooth muscle actin (α -Sma) shows the expression of Snail1 both in renal epithelial and stromal cells of fibrotic kidney. (Bottom) IF for Prrx1 and α -Sma shows the expression of Prrx1 only in stromal cells of fibrotic kidney. (c) Expression mature mouse *miR-15* family members in fibrotic kidney shows the upregulation of several members. SHAM kidneys were used as control.

SHAM: kidney of mice that had been subjected to surgery but not UUO, used as control. Plots and bars represent 3 independent experiments from 3 animals as biological replicates, and asterisks indicate significant p-value in t-test to the control (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$).

Chapter 5

DISCUSSION

Previous findings in the lab were pointing at a complex regulatory network between EMT-TFs, including positive and negative correlations. *Prrx1*-induced EMT was different from that induced by other classical EMT-TFs, in particular *Snail1* (Ocana et al., 2012). *Snail1* is a strong epithelial repressor and *Prrx1* is a better mesenchymal inducer, and they differ in their ability to induce stem cell properties (Nieto et al., 2016). Here, we provide evidence of the existence of a Gene Regulatory Network (GRN) in which *Prrx1*, through the activation of the miR-15 family, attenuates the expression of *Snail1*. In turn, *Snail1* represses *Prrx1* transcription, resulting in complementary/mutually exclusive expression patterns both during vertebrate embryonic development and also in diseases such as cancer and fibrosis, and indicating that these two transcription factors not only trigger but also promote their own mode of EMT. This GRN also leads to a temporal control in embryos and cell lines, where *Snail1* expression precedes that of *Prrx1*. In addition, our data suggest that *Prrx1* could have an instrumental role in promoting the progression towards a full EMT, with important implications in embryonic development and particularly, in cancer.

5.1 The GRN that controls EMT programs

Epithelial cells transition to a variety of mesenchymal states to fulfil different roles in different contexts. The existence of a variety of mesenchymal cells allows for a high degree of cell heterogeneity in terms of potential fates and therefore, functions. They are all called mesenchymal based on morphology, the expression of particular markers (although there is no universal marker to define them), and very often, their ability to migrate. Differences arise from the tissue of origin and the signals that they receive from the microenvironment. These phenotypic transitions are governed by extracellular signals that activate a plethora of the so-called epithelial-to mesenchymal transcription factors (EMT-TFs).

Functions of the EMT-TFs are not limited to trigger EMT in epithelial cells but also to participate in regulatory networks to maintain the mesenchymal state. Different mesenchymal cells, both of embryonic, healthy adult or pathological origin, express different combinations of EMT-TFs, known as EMT-TF code (Nieto, 2017), which determine the overall state, function and behavior of the cell. For instance, some EMT-TFs combinations may endow cells with the ability to migrate in a collective manner,

while others can lead to single cell migration. In general, when associated with migratory programs, mesenchymal states coincide with low proliferation, which will allow a more efficient migration (Nieto et al., 2016). Studying the differences between all these EMT-TFs is important for our understanding of this complexity during embryonic development, which can ultimately help to distinguish the key altered cellular and molecular mechanisms in disease. Thus, we have focused our study on the crosstalk of two dissimilar EMT-TFs; Snail1 and Prrx1.

The EMT-TF code implies that several EMT-TFs are co-expressed in cell populations during embryonic development and adult physio-pathology. However, previous preliminary data from the lab and this work confirms that Snail1 and Prrx1 are expressed in a complementary manner during vertebrate development, in cancer and in fibrosis ([Figures 12-14](#)) (Ocana et al., 2012). Snail1 and Snail2 are also expressed in complementary patterns in certain regions and, when analyzed in different species like mouse or chicken, their expression patterns are inverted. However, they have similar functions in the same cell population, like the case of PNC cells, where their expression is shuffled from one species to another during vertebrate evolution and one family member can substitute for the function of the other if expressed in the appropriate territory (Locascio et al., 2002).

Interestingly, Snail1 and Prrx1, which belong to two different genes families, can also fulfill similar functions in different species, including instructing heart laterality (Ocana et al., 2017). As such, Prrx1 in the fish and Snail1 in the mouse displace the posterior pole of the heart to the left. This was tinkered during evolution and fixed to fulfill the species-specific requirements imposed by topology and tissue stiffness through the use of tissue-specific enhancers. Thus, only Snail1 is expressed in the mouse in the appropriate territories to drive heart laterality, and the same is true for Prrx1 in the fish. However, it is worth noting here that within one species/individual, Snail1 and Prrx1 are still expressed in mutually exclusive domains, strongly suggesting that in addition to the tight control imposed by the tissue-specific enhancers, there might be a superimposed mechanism that refines this information through the establishment of a mutual regulatory mechanism. In this work, we describe such a mechanism. We find that Snail1 and Prrx1 behave as mutual repressors. Interestingly, they use different mechanisms. Albeit in some contexts Snail1 can function as an activator (Hsu et al., 2014), in the vast majority of cases, it is a very strong transcriptional repressor (Cano et al., 2000), and this is the mechanism that it uses to repress Prrx1 expression. Prrx1,

in turn, is a transcriptional activator (Norris and Kern, 2001, Grueneberg et al., 1992) and accordingly, here we show that it represses Snail1 indirectly through the activation of a repressor, in this case the miR-15 family. This also provides a certain degree of variability in the repression. As such, microRNAs are transcriptional attenuators, and as miR-15 is a family of microRNAs, one or several family members can be used in specific cell contexts to impinge a particular level of repression.

Following the developmental time in a particular tissue, it is possible to infer potential regulatory scenarios in which the mutual repression of these TFs could be integrated. For instance, in the mouse, Snail1 is expressed in the premigratory neural crest and Prrx1 is expressed in subpopulations of migratory crest (Figure 13). During somite formation, Snail1 is expressed in the precursors of the whole somite and, later on in the mature somites, it is expressed in the sclerotome and Prrx1 in the dermomyotome, also showing complementary expression (Figures 12-13). This also implies that Prrx1 is activated in some descendants of Snail1 expressing cells. Altogether, this points to the successive upregulation of EMT-TFs during the development of the neural crest and the somites, providing a hierarchical temporal order in the activation of these two transcription factors. As such, we provide evidence of this temporal hierarchy in cell culture experiments, with Snail1 being activated at earlier time points in response to an EMT-inducing signal (Figure 25).

Interestingly, the inducing signal is the same for Snail1 and Prrx1 (the TGF- β superfamily; Ocaña et al., 2012 and this work), and thus, the question is why they are not simultaneously activated when the signal is available in the embryo or in cancer cells. One explanation is that Snail1 bears a poised promoter (Lagha et al., 2013), explaining why it usually is a very early response gene activated shortly after TGF- β administration (Peinado et al., 2003, Valdes et al., 2002). Snail1 fast activation may set an inhibitory scenario for Prrx1 upregulation, scenario that may change when Snail1 first activation peak starts to decline and/or when the duration of the TGF- β signal accumulates sufficient stimulus to induce Prrx1. This establishes a temporal order of predominant EMT programs, firstly that mediated by Snail1, a strong epithelial repressor needed for cells to detach from their neighbors and subsequently, a strong mesenchymal inducer, Prrx1. Prrx1 will, in turn, promote the acquisition of robust mesenchymal features and the maintenance of this mesenchymal phenotype for cells to migrate to their final destination. Needless to say, the cells do not only express these two transcription factors but also members of the other EMT-TF families in a context-

dependent combination that will determine their final phenotype. Nonetheless, Snail1 and Prrx1, as already discussed, are mutually exclusive, thereby promoting their own mode of EMT. Both are usually accompanied by other EMT-TFs, including Snail2 and Twist (unpublished data from the lab). We also found in this work that Prrx1 further reinforces its own program by activating its own promoter (Figure 16). Interestingly, Snail1 represses its promoter, thereby fine-tuning and modulating its own expression (Peiro et al., 2006). This can also contribute to the establishment of the temporal hierarchy of EMT-TF activation, promoting Prrx1 upregulation, as we observe during embryonic development. Altogether, these mutual regulations and the response to the inducing signal establish a Gene Regulatory Network (GRN) that control the EMT programs as summarized in Figure 27.

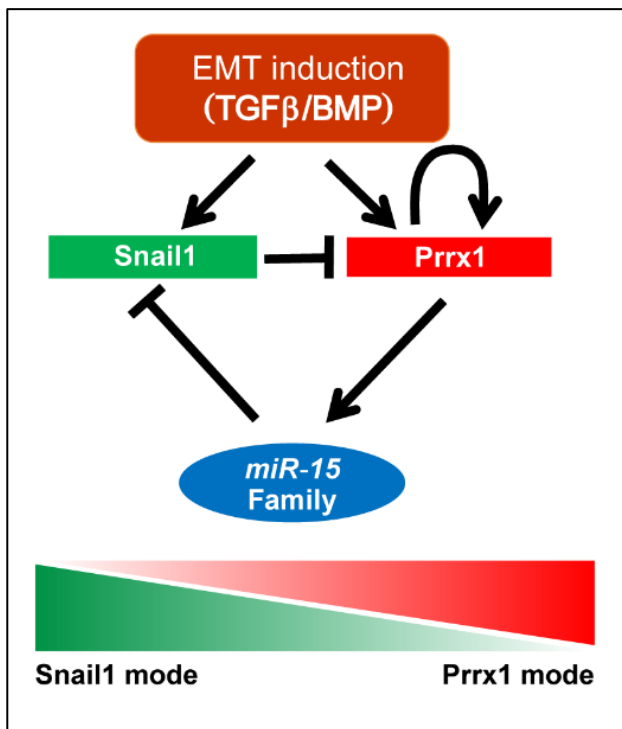


Figure 27. Working model for a gene regulatory network in which miR15 family may coordinate the transition from Snail1- to Prrx1-mediated EMT programs.

5.2 Differences and similarities in Snail1 and Prrx1-induced Programs

Snail1 and Prrx1 EMT programs differ in some important aspects. As such, EMT has been associated with the acquisition of stem cell properties (Mani et al., 2008, Morel et al., 2008) but Prrx1 represses stemness (Ocana et al., 2012). This has been associated with the plasticity required in both embryonic and cancer cells to revert to a more epithelial phenotype at the site of their final destination to differentiate into

different organs during development or macrometastases during cancer progression (see (Ocana et al., 2012, Tsai et al., 2012, Celia-Terrassa et al., 2012, Beerling et al., 2016). As such, Prrx1 needs to be downregulated in cancer cells to promote the reversion, and its downregulation implies an increase in stem cell properties that are also essential for metastatic growth. This also explains why there is a positive correlation between high levels of Prrx1 and better prognosis in cancer patients (Ocaña et al., 2012). Those cells, may be stuck at a very mesenchymal state, lacking the plasticity required to colonize and to restart proliferation in distant organs. In line with that, we have found that miR-15 family members are direct targets of Prrx1 and also that they correlate with better survival rate in patients, where they are expressed at high levels.

Importantly, the relationship between EMT and stemness has been mainly studied in adult mammary gland and in cancer cells, but these differences in conferring stem cell properties may also have a correlate during embryonic development. Snail1 expressing cells might be in more stem state, as they can give rise to cells with a variety of fates. For instance, Snail1 is expressed early in primitive streak and the premigratory neural crest, both of which later give rise to a huge variety of cell types (Thiery et al., 2009). As such, it has been proposed that the premigratory neural crest may still retain characteristic of the pluripotent blastula stage (Buitrago-Delgado et al., 2015). These pluripotent territories do not express Prrx1, which is rather expressed in cells that have been already committed to particular lineages such as a specific mesodermal or crest subpopulations. This is also compatible with the already discussed temporal hierarchy in the activation of these two EMT-TFs that occurs in parallel with commitment to different fates.

As much as Snail1 and Prrx1 trigger different EMT programs with respect to stemness and association with survival of cancer patients (Ocaña et al, 2012; see below), they have important common functional roles inducing the transition towards a mesenchymal state, activating migration and invasion. Snail1 induces migration and invasion in many different contexts through activation of a variety of factors including matrix metalloproteinases (MMPs) that break collagens and enable cells to migrate through basal lamina (Miyoshi et al., 2004, De Craene et al., 2005, Rowe et al., 2009, Yokoyama et al., 2003) and the attenuation of cell proliferation (Vega et al., 2004). Prrx1 also decreases cell proliferation and induces invasiveness (Ocana et al., 2012, Sugiyama et al., 2015, Guo et al., 2015). These roles of Prrx1 are likely, at least in part,

to be mediated by miR-15, as some family members (miR-424 and miR-503) induce migration and invasion (Li et al., 2014a, Drasin et al., 2015b).

5.3 Prrx1 expression, a switch from partial to full EMT

As previously discussed, Snail1 is an essential EMT initiator as an epithelial repressor, while Twist1 and Prrx1 are much better mesenchymal inducers (Tran et al., 2011, Peinado et al., 2007, Nieto et al., 2016). Therefore, and again in agreement with the temporal progression towards the mesenchymal phenotype, epithelial repression will be required at early stages to promote the downregulation of cell-cell adhesion complexes and the subsequent acquisition of mesenchymal characteristics including the secretion of metalloproteinases will help break the basement membrane and the delamination process. Together with cytoskeletal reorganizations typical of mesenchymal cells, they will be also able to delaminate and start migrating. Then, cells usually need to migrate long distances and maintain the mesenchymal phenotype, for which Prrx1 seems to be essential in different contexts. Thus, we propose a sequential program as shown in Figure 28, compatible with our analysis in cultured cells and that has a correlate in the expression and behavior of Snail1 and Prrx1 in embryos, kidney fibrosis and cancer.

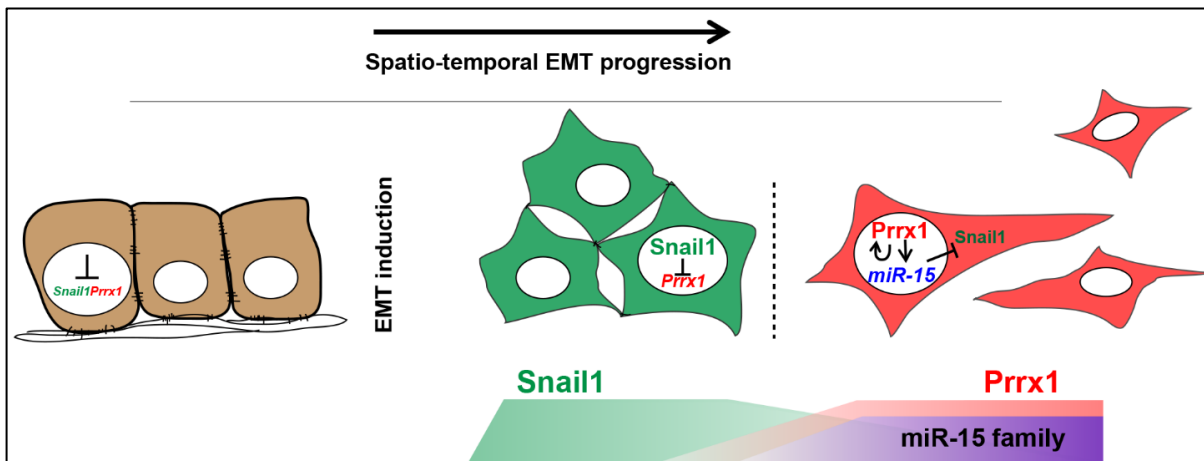


Figure 28. Hypothetical spatiotemporal EMT progression. Epithelial cells that tightly adhere to each other and to the basement membrane, partially lose their contact upon EMT induction and Snail1 expression. Prrx1 is upregulated and induces miR-15 family members that in turn, attenuate Snail1 expression, and cells switch to a Prrx1-mediated program. High levels of Prrx1 make cells progress to a full EMT state.

The gradual upregulation of Prrx1 that we see associated with the migration of embryonic cells and the progression of the mesenchymal state in cancer cells, fits with our analysis of the described subpopulations of mouse skin tumor cells (Pastushenko et al., 2018). Indeed, this mouse model reveals an increase in mesenchymal character concomitant with a gradual increase in the expression of Prrx1. In line with findings of Ocaña and colleagues (Ocana et al., 2012), the subpopulation with intermediate EMT state and lowest expression level of Prrx1 showed the highest potential for metastatic colonization (Pastushenko et al., 2018). This finding, supports the idea that Prrx1 expression is increased during EMT progression, and regulates the transition from partial to full EMT. In fibrosis, epithelial cells undergo a partial EMT, concomitant with the activation of Snail1, but not Prrx1, again supporting the hypothesis that Prrx1 activation is required to acquire a full EMT phenotype. However, we still need to prove this hypothesis *in vivo*. A very appropriate model would be a transgenic mouse model in which we induce fibrosis as before (Grande et al., 2015), and simultaneously ectopically express Prrx1 in the renal epithelial cells that otherwise we know that fail to activate it and that only undergo a partial-EMT. The prediction is that renal epithelial cells will now progress in the mesenchymal state and delaminate from the tubules.

In summary, a full EMT phenotype is incompatible with metastatic colonization and would very likely induce the disruption of renal tubules in fibrosis. All this has important implications in both diseases. Unpublished data from the lab indicate that the EMT may be subjected to attenuation during evolution, suggesting that the partial EMT state is favored. Importantly, the partial EMT favors collective migration in development and cancer and with that, plasticity to revert and therefore, colonization of embryonic cells at sites of destination and of cancer cells at distant organs, respectively. It also favors wound healing, where cells that undergo partial EMT have to re-epithelialize to fully close the wound and reacquire epithelial homeostasis. If the EMT phenotype has been attenuated through evolution, favoring the partial EMT, this implies that it will favor robust embryonic development and improved wound healing (including reversion of fibrosis), but also metastasis. The whole concept makes sense in terms of evolutionary fitness, as conditions that favor embryonic development and wound healing are likely subjected to high evolutionary pressure. With that, promoting cancer metastases would appear as a tradeoff. Needless to say, new animal models out of the scope of this work are required to address this relevant biological question.

Overall, in this work, we describe a GRN that operates in development and in disease to regulate the activation of different EMT programs, providing a superimposed degree of controlled heterogeneity required for the EMT, a crucial and complex biological process.

CONCLUSIONS

Conclusions

- We have defined a Gene Regulatory Network that controls EMT programs. This GRN is characterized by:
 - A complementary expression of *Prrx1* and *Snail1* that results from a double negative feedback loop, in which *Snail1* directly represses the transcription of *Prrx1* and in turn, *Prrx1* attenuates *Snail1* expression through direct activation of miR-15 family members.
 - *Prrx1* induces its own transcription in order to promote its own EMT mode.
 - The integration of the inducing signal (TGF-beta) with this GRN leads to the sequential activity of *Snail1* and *Prrx1* and, therefore, of their associated EMT programs
- This GRN is conserved during vertebrate development and in pathological states including cancer and fibrosis
- High expression of *Prrx1* may trigger the transition from partial EMT to full EMT, with implications in development and disease, as
 - High *PRRX1* is a sign of better prognosis in cancer: full EMT, no reversion to a more epithelial state, no metastatic colonization
 - The absence of *Prrx1* allows reversibility in renal fibrosis: partial EMT, potentiality to revert, no tubular disruption

REFERENCES

- ACETO, N., BARDIA, A., MIYAMOTO, D. T., DONALDSON, M. C., WITTNER, B. S., SPENCER, J. A., YU, M., PELY, A., ENGSTROM, A., ZHU, H., BRANNIGAN, B. W., KAPUR, R., STOTT, S. L., SHIODA, T., RAMASWAMY, S., TING, D. T., LIN, C. P., TONER, M., HABER, D. A. & MAHESWARAN, S. 2014. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*, 158, 1110-1122.
- ACLOQUE, H., ADAMS, M. S., FISHWICK, K., BRONNER-FRASER, M. & NIETO, M. A. 2009. Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. *The Journal of clinical investigation*, 119, 1438-49.
- ACLOQUE, H., OCANA, O. H., MATHEU, A., RIZZOTI, K., WISE, C., LOVELL-BADGE, R. & NIETO, M. A. 2011. Reciprocal repression between Sox3 and snail transcription factors defines embryonic territories at gastrulation. *Dev Cell*, 21, 546-58.
- ACLOQUE, H., OCANA, O. H. & NIETO, M. A. 2012. Mutual exclusion of transcription factors and cell behaviour in the definition of vertebrate embryonic territories. *Curr Opin Genet Dev*, 22, 308-14.
- AHLSTROM, J. D. & ERICKSON, C. A. 2009. The neural crest epithelial-mesenchymal transition in 4D: a 'tail' of multiple non-obligatory cellular mechanisms. *Development*, 136, 1801-12.
- AIELLO, N. M., MADDIPATI, R., NORGARD, R. J., BALLI, D., LI, J., YUAN, S., YAMAZOE, T., BLACK, T., SAHMOUD, A., FURTH, E. E., BAR-SAGI, D. & STANGER, B. Z. 2018. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Developmental cell*, 45, 681-695 e4.
- ARNOUX, V., NASSOUR, M., L'HELGOUALC'H, A., HIPSKIND, R. A. & SAVAGNER, P. 2008. Erk5 controls Slug expression and keratinocyte activation during wound healing. *Molecular biology of the cell*, 19, 4738-49.
- BARRALLO-GIMENO, A. & NIETO, M. A. 2005. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development*, 132, 3151-61.
- BARTEL, D. P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215-33.
- BARTEL, D. P. 2018. Metazoan MicroRNAs. *Cell*, 173, 20-51.
- BATLLE, E., SANCHO, E., FRANCI, C., DOMINGUEZ, D., MONFAR, M., BAULIDA, J. & GARCIA DE HERREROS, A. 2000. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol*, 2, 84-9.
- BEERLING, E., SEINSTRA, D., DE WIT, E., KESTER, L., VAN DER VELDEN, D., MAYNARD, C., SCHAFER, R., VAN DIEST, P., VOEST, E., VAN OUDENAARDEN, A., VRISEKOO, N. & VAN RHEENEN, J. 2016. Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep*, 14, 2281-8.
- BEHRENS, J., VON KRIES, J. P., KUHL, M., BRUHN, L., WEDLICH, D., GROSSCHEDL, R. & BIRCHMEIER, W. 1996. Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature*, 382, 638-42.

- BOUTET, A., DE FRUTOS, C. A., MAXWELL, P. H., MAYOL, M. J., ROMERO, J. & NIETO, M. A. 2006. Snail activation disrupts tissue homeostasis and induces fibrosis in the adult kidney. *The EMBO journal*, 25, 5603-13.
- BRACKEN, C. P., GREGORY, P. A., KOLESNIKOFF, N., BERT, A. G., WANG, J., SHANNON, M. F. & GOODALL, G. J. 2008. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res*, 68, 7846-54.
- BUITRAGO-DELGADO, E., NORDIN, K., RAO, A., GEARY, L. & LABONNE, C. 2015. NEURODEVELOPMENT. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science*, 348, 1332-5.
- BURK, U., SCHUBERT, J., WELLNER, U., SCHMALHOFER, O., VINCAN, E., SPADERNA, S. & BRABLETZ, T. 2008. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep*, 9, 582-9.
- CAJA, L., BERTRAN, E., CAMPBELL, J., FAUSTO, N. & FABREGAT, I. 2011. The transforming growth factor-beta (TGF-beta) mediates acquisition of a mesenchymal stem cell-like phenotype in human liver cells. *J Cell Physiol*, 226, 1214-23.
- CALIN, G. A., DUMITRU, C. D., SHIMIZU, M., BICHI, R., ZUPO, S., NOCH, E., ALDLER, H., RATTAN, S., KEATING, M., RAI, K., RASSENTI, L., KIPPS, T., NEGRINI, M., BULLRICH, F. & CROCE, C. M. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 15524-9.
- CANO, A. & NIETO, M. A. 2008. Non-coding RNAs take centre stage in epithelial-to-mesenchymal transition. *Trends Cell Biol*, 18, 357-9.
- CANO, A., PEREZ-MORENO, M. A., RODRIGO, I., LOCASCIO, A., BLANCO, M. J., DEL BARRIO, M. G., PORTILLO, F. & NIETO, M. A. 2000. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature cell biology*, 2, 76-83.
- CARVER, E. A., JIANG, R., LAN, Y., ORAM, K. F. & GRIDLEY, T. 2001. The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Molecular and cellular biology*, 21, 8184-8.
- CELIA-TERRASSA, T., MECA-CORTES, O., MATEO, F., MARTINEZ DE PAZ, A., RUBIO, N., ARNAL-ESTAPE, A., ELL, B. J., BERMUDO, R., DIAZ, A., GUERRA-REBOLLO, M., LOZANO, J. J., ESTARAS, C., ULLOA, C., ALVAREZ-SIMON, D., MILA, J., VILELLA, R., PACIUCCI, R., MARTINEZ-BALBAS, M., DE HERREROS, A. G., GOMIS, R. R., KANG, Y., BLANCO, J., FERNANDEZ, P. L. & THOMSON, T. M. 2012. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J Clin Invest*, 122, 1849-68.
- CHUNG, W., EUM, H. H., LEE, H. O., LEE, K. M., LEE, H. B., KIM, K. T., RYU, H. S., KIM, S., LEE, J. E., PARK, Y. H., KAN, Z., HAN, W. & PARK, W. Y. 2017. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature communications*, 8, 15081.
- CIRUNA, B. & ROSSANT, J. 2001. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Developmental cell*, 1, 37-49.

- DE CRAENE, B. & BERX, G. 2013. Regulatory networks defining EMT during cancer initiation and progression. *Nature reviews. Cancer*, 13, 97-110.
- DE CRAENE, B., GILBERT, B., STOVE, C., BRUYNEEL, E., VAN ROY, F. & BERX, G. 2005. The transcription factor snail induces tumor cell invasion through modulation of the epithelial cell differentiation program. *Cancer research*, 65, 6237-44.
- DEL BARRIO, M. G. & NIETO, M. A. 2004. Relative expression of Slug, RhoB, and HNK-1 in the cranial neural crest of the early chicken embryo. *Dev Dyn*, 229, 136-9.
- DIAZ-LOPEZ, A., MORENO-BUENO, G. & CANO, A. 2014. Role of microRNA in epithelial to mesenchymal transition and metastasis and clinical perspectives. *Cancer management and research*, 6, 205-16.
- DONG, J., HU, Y., FAN, X., WU, X., MAO, Y., HU, B., GUO, H., WEN, L. & TANG, F. 2018. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome biology*, 19, 31.
- DRASIN, D. J., GUARNIERI, A. L., NEELAKANTAN, D., KIM, J., CABRERA, J. H., WANG, C.-A., ZABEREZHNYI, V., GASPARINI, P., CASCIONE, L. & HUEBNER, K. 2015a. TWIST1-induced miR-424 reversibly drives mesenchymal programming while inhibiting tumor initiation. *Cancer research*, 75, 1908-1921.
- DRASIN, D. J., GUARNIERI, A. L., NEELAKANTAN, D., KIM, J., CABRERA, J. H., WANG, C. A., ZABEREZHNYI, V., GASPARINI, P., CASCIONE, L., HUEBNER, K., TAN, A. C. & FORD, H. L. 2015b. TWIST1-Induced miR-424 Reversibly Drives Mesenchymal Programming while Inhibiting Tumor Initiation. *Cancer research*, 75, 1908-21.
- EBERT, M. S., NEILSON, J. R. & SHARP, P. A. 2007. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature methods*, 4, 721-6.
- FABREGAT, I., Malfettone, A. & Soukupo, J. 2016. New Insights into the Crossroads between EMT and Stemness in the Context of Cancer. *J Clin Med*, 5.
- FISCHER, K. R., DURRANS, A., LEE, S., SHENG, J., LI, F., WONG, S. T., CHOI, H., EL RAYES, T., RYU, S., TROEGER, J., SCHWABE, R. F., VAHDAT, L. T., ALTORKI, N. K., MITTAL, V. & GAO, D. 2015. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*, 527, 472-6.
- FRANCO, D. L., MAINEZ, J., VEGA, S., SANCHO, P., MURILLO, M. M., DE FRUTOS, C. A., DEL CASTILLO, G., LOPEZ-BLAU, C., FABREGAT, I. & NIETO, M. A. 2010. Snail1 suppresses TGF-beta-induced apoptosis and is sufficient to trigger EMT in hepatocytes. *Journal of cell science*, 123, 3467-77.
- FROMM, B., BILLIPP, T., PECK, L. E., JOHANSEN, M., TARVER, J. E., KING, B. L., NEWCOMB, J. M., SEMPERE, L. F., FLATMARK, K., HOVIG, E. & PETERSON, K. J. 2015. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annual review of genetics*, 49, 213-42.
- GENTLEMAN, R. C., CAREY, V. J., BATES, D. M., BOLSTAD, B., DETTLING, M., DUDOIT, S., ELLIS, B., GAUTIER, L., GE, Y., GENTRY, J., HORNIK, K., HOTHORN, T., HUBER, W., IACUS, S., IRIZARRY, R., LEISCH, F., LI, C., MAECHLER, M., ROSSINI, A. J., SAWITZKI, G., SMITH, C., SMYTH, G., TIERNEY, L., YANG, J. Y. & ZHANG, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5, R80.

- GILLES, C., POLETTE, M., MESTDAGT, M., NAWROCKI-RABY, B., RUGGERI, P., BIREMBAUT, P. & FOIDART, J. M. 2003. Transactivation of vimentin by beta-catenin in human breast cancer cells. *Cancer research*, 63, 2658-64.
- GRANDE, M. T., SANCHEZ-LAORDEN, B., LOPEZ-BLAU, C., DE FRUTOS, C. A., BOUTET, A., AREVALO, M., ROWE, R. G., WEISS, S. J., LOPEZ-NOVOA, J. M. & NIETO, M. A. 2015. Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nature medicine*, 21, 989-97.
- GREGORY, P. A., BERT, A. G., PATERSON, E. L., BARRY, S. C., TSYKIN, A., FARSHID, G., VADAS, M. A., KHEW-GOODALL, Y. & GOODALL, G. J. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol*, 10, 593-601.
- GRUENEBERG, D. A., NATESAN, S., ALEXANDRE, C. & GILMAN, M. Z. 1992. Human and Drosophila homeodomain proteins that enhance the DNA-binding activity of serum response factor. *Science*, 257, 1089-95.
- GUO, J., FU, Z., WEI, J., LU, W., FENG, J. & ZHANG, S. 2015. PRRX1 promotes epithelial-mesenchymal transition through the Wnt/beta-catenin pathway in gastric cancer. *Medical oncology*, 32, 393.
- GYORFFY, B., LANCZKY, A., EKLUND, A. C., DENKERT, C., BUDCZIES, J., LI, Q. & SZALLASI, Z. 2010. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*, 123, 725-31.
- HAMBURGER, V. & HAMILTON, H. L. 1951. A series of normal stages in the development of the chick embryo. *Journal of morphology*, 88, 49-92.
- HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell*, 144, 646-74.
- HARPER, K. L., SOSA, M. S., ENTENBERG, D., HOSSEINI, H., CHEUNG, J. F., NOBRE, R., AVIVAR-VALDERAS, A., NAGI, C., GIRNIUS, N., DAVIS, R. J., FARIAS, E. F., CONDEELIS, J., KLEIN, C. A. & AGUIRRE-GHISO, J. A. 2016. Mechanism of early dissemination and metastasis in Her2(+) mammary cancer. *Nature*.
- HOSSEINI, H., OBRADOVIC, M. M., HOFFMANN, M., HARPER, K. L., SOSA, M. S., WERNER-KLEIN, M., NANDURI, L. K., WERNO, C., EHRL, C., MANECK, M., PATWARY, N., HAUNSCHILD, G., GUZVIC, M., REIMELT, C., GRAUVOGL, M., EICHNER, N., WEBER, F., HARTKOPF, A. D., TARAN, F. A., BRUCKER, S. Y., FEHM, T., RACK, B., BUCHHOLZ, S., SPANG, R., MEISTER, G., AGUIRRE-GHISO, J. A. & KLEIN, C. A. 2016. Early dissemination seeds metastasis in breast cancer. *Nature*.
- HSU, D. S., WANG, H. J., TAI, S. K., CHOU, C. H., HSIEH, C. H., CHIU, P. H., CHEN, N. J. & YANG, M. H. 2014. Acetylation of snail modulates the cytokinome of cancer cells to enhance the recruitment of macrophages. *Cancer Cell*, 26, 534-48.
- HUTCHINSON, J., FOGARTY, A., HUBBARD, R. & MCKEEVER, T. 2015. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *The European respiratory journal*, 46, 795-806.

- IHIDA-STANSBURY, K., MCKEAN, D. M., GEBB, S. A., MARTIN, J. F., STEVENS, T., NEMENOFF, R., AKESON, A., VAUGHN, J. & JONES, P. L. 2004. Paired-related homeobox gene Prx1 is required for pulmonary vascular development. *Circulation research*, 94, 1507-14.
- IMAI, T., HORIUCHI, A., WANG, C., OKA, K., OHIRA, S., NIKAIDO, T. & KONISHI, I. 2003. Hypoxia attenuates the expression of E-cadherin via up-regulation of SNAIL in ovarian carcinoma cells. *The American journal of pathology*, 163, 1437-47.
- IRIZARRY, R. A., BOLSTAD, B. M., COLLIN, F., COPE, L. M., HOBBS, B. & SPEED, T. P. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research*, 31, e15.
- JAILLON, O., AURY, J. M., BRUNET, F., PETIT, J. L., STANGE-THOMANN, N., MAUCELI, E., BOUNEAU, L., FISCHER, C., OZOUF-COSTAZ, C., BERNOT, A., NICAUD, S., JAFFE, D., FISHER, S., LUTFALLA, G., DOSSAT, C., SEGURENS, B., DASILVA, C., SALANOUBAT, M., LEVY, M., BOUDET, N., CASTELLANO, S., ANTHOUARD, V., JUBIN, C., CASTELLI, V., KATINKA, M., VACHERIE, B., BIEMONT, C., SKALLI, Z., CATTOLICO, L., POULAIN, J., DE BERARDINIS, V., CRUAUD, C., DUPRAT, S., BROTTIER, P., COUTANCEAU, J. P., GOUZY, J., PARRA, G., LARDIER, G., CHAPPLE, C., MCKERNAN, K. J., MCEWAN, P., BOSAK, S., KELLIS, M., VOLFF, J. N., GUIGO, R., ZODY, M. C., MESIROV, J., LINDBLAD-TOH, K., BIRREN, B., NUSBAUM, C., KAHN, D., ROBINSON-RECHAVI, M., LAUDET, V., SCHACHTER, V., QUETIER, F., SAURIN, W., SCARPELLI, C., WINCKER, P., LANDER, E. S., WEISSENBAACH, J. & ROEST CROLLIUS, H. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431, 946-57.
- KIDA, Y., ASAHINA, K., TERAOKA, H., GITELMAN, I. & SATO, T. 2007. Twist relates to tubular epithelial-mesenchymal transition and interstitial fibrogenesis in the obstructed kidney. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 55, 661-73.
- KIM, J. H., LEE, S. R., LI, L. H., PARK, H. J., PARK, J. H., LEE, K. Y., KIM, M. K., SHIN, B. A. & CHOI, S. Y. 2011a. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one*, 6, e18556.
- KIM, N. H., KIM, H. S., LI, X. Y., LEE, I., CHOI, H. S., KANG, S. E., CHA, S. Y., RYU, J. K., YOON, D., FEARON, E. R., ROWE, R. G., LEE, S., MAHER, C. A., WEISS, S. J. & YOOK, J. I. 2011b. A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. *J Cell Biol*, 195, 417-33.
- KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B. & SCHILLING, T. F. 1995. Stages of embryonic development of the zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists*, 203, 253-310.
- KLIJN, C., DURINCK, S., STAWISKI, E. W., HAVERTY, P. M., JIANG, Z., LIU, H., DEGENHARDT, J., MAYBA, O., GNAD, F., LIU, J., PAU, G., REEDER, J., CAO, Y., MUKHYALA, K., SELVARAJ, S. K., YU, M., ZYNDA, G. J., BRAUER, M. J., WU, T. D., GENTLEMAN, R. C., MANNING, G., YAUCH, R. L., BOURGON, R., STOKOE, D., MODRUSAN, Z., NEVE, R. M., DE SAUVAGE, F. J., SETTLEMAN, J., SESHAGIRI, S. & ZHANG, Z. 2015. A comprehensive transcriptional portrait of human cancer cell lines. *Nature biotechnology*, 33, 306-12.

- KNOFLER, M. & POLLHEIMER, J. 2013. Human placental trophoblast invasion and differentiation: a particular focus on Wnt signaling. *Frontiers in genetics*, 4, 190.
- KREBS, A. M., MITSCHKE, J., LASIERRA LOSADA, M., SCHMALHOFER, O., BOERRIES, M., BUSCH, H., BOETTCHER, M., MOUGIAKAKOS, D., REICHARDT, W., BRONSERT, P., BRUNTON, V. G., PILARSKY, C., WINKLER, T. H., BRABLETZ, S., STEMMLER, M. P. & BRABLETZ, T. 2017. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nature cell biology*, 19, 518-529.
- KRUTHOF-DE JULIO, M., ALVAREZ, M. J., GALLI, A., CHU, J., PRICE, S. M., CALIFANO, A. & SHEN, M. M. 2011. Regulation of extra-embryonic endoderm stem cell differentiation by Nodal and Cripto signaling. *Development*, 138, 3885-95.
- LAGHA, M., BOTHMA, J. P., ESPOSITO, E., NG, S., STEFANIK, L., TSUI, C., JOHNSTON, J., CHEN, K., GILMOUR, D. S., ZEITLINGER, J. & LEVINE, M. S. 2013. Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell*, 153, 976-87.
- LAMBERT, A. W., PATTABIRAMAN, D. R. & WEINBERG, R. A. 2017. Emerging Biological Principles of Metastasis. *Cell*, 168, 670-691.
- LANCZKY, A., NAGY, A., BOTTAI, G., MUNKACSY, G., SZABO, A., SANTARPIA, L. & GYORFFY, B. 2016. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast cancer research and treatment*, 160, 439-446.
- LATIL, M., NASSAR, D., BECK, B., BOUMAHDI, S., WANG, L., BRISEBARRE, A., DUBOIS, C., NKUSI, E., LENGLEZ, S., CHECINSKA, A., VERCAUTEREN DRUBBEL, A., DEVOS, M., DECLERCQ, W., YI, R. & BLANPAIN, C. 2017. Cell-Type-Specific Chromatin States Differentially Prime Squamous Cell Carcinoma Tumor-Initiating Cells for Epithelial to Mesenchymal Transition. *Cell stem cell*, 20, 191-204 e5.
- LI, Y., LI, W., YING, Z., TIAN, H., ZHU, X., LI, J. & LI, M. 2014a. Metastatic heterogeneity of breast cancer cells is associated with expression of a heterogeneous TGFβ-activating miR424-503 gene cluster. *Cancer research*, 74, 6107-18.
- LI, Y., LI, W., YING, Z., TIAN, H., ZHU, X., LI, J. & LI, M. 2014b. Metastatic Heterogeneity of Breast Cancer Cells Is Associated with Expression of a Heterogeneous TGFβ-Activating miR424–503 Gene Cluster. *Cancer research*, 74, 6107-6118.
- LOCASCIO, A., MANZANARES, M., BLANCO, M. J. & NIETO, M. A. 2002. Modularity and reshuffling of Snail and Slug expression during vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 16841-6.
- LOVISA, S., LEBLEU, V. S., TAMPE, B., SUGIMOTO, H., VADNAGARA, K., CARSTENS, J. L., WU, C. C., HAGOS, Y., BURCKHARDT, B. C., PENTCHEVA-HOANG, T., NISCHAL, H., ALLISON, J. P., ZEISBERG, M. & KALLURI, R. 2015. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. *Nature medicine*, 21, 998-1009.
- LU, M. F., CHENG, H. T., KERN, M. J., POTTER, S. S., TRAN, B., DIEKWISCH, T. G. & MARTIN, J. F. 1999. prx-1 functions cooperatively with another paired-related homeobox gene, prx-2, to maintain cell fates within the craniofacial mesenchyme. *Development*, 126, 495-504.

- LU, X. & KANG, Y. 2010. Hypoxia and hypoxia-inducible factors: master regulators of metastasis. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 16, 5928-35.
- MANI, S. A., GUO, W., LIAO, M. J., EATON, E. N., AYYANAN, A., ZHOU, A. Y., BROOKS, M., REINHARD, F., ZHANG, C. C., SHIPITSIN, M., CAMPBELL, L. L., POLYAK, K., BRISKEN, C., YANG, J. & WEINBERG, R. A. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133, 704-15.
- MANZANARES, M., LOCASCIO, A. & NIETO, M. A. 2001. The increasing complexity of the Snail gene superfamily in metazoan evolution. *Trends Genet*, 17, 178-81.
- MARTIK, M. L. & BRONNER, M. E. 2017. Regulatory Logic Underlying Diversification of the Neural Crest. *Trends Genet*, 33, 715-727.
- MARTIN, J. F., BRADLEY, A. & OLSON, E. N. 1995. The paired-like homeo box gene MHOX is required for early events of skeletogenesis in multiple lineages. *Genes & development*, 9, 1237-49.
- MCKEAN, D. M., SISBARRO, L., ILIC, D., KAPLAN-ALBURQUERQUE, N., NEMENOFF, R., WEISER-EVANS, M., KERN, M. J. & JONES, P. L. 2003. FAK induces expression of Prx1 to promote tenascin-C-dependent fibroblast migration. *The Journal of cell biology*, 161, 393-402.
- MIYOSHI, A., KITAJIMA, Y., SUMI, K., SATO, K., HAGIWARA, A., KOGA, Y. & MIYAZAKI, K. 2004. Snail and SIP1 increase cancer invasion by upregulating MMP family in hepatocellular carcinoma cells. *British journal of cancer*, 90, 1265-73.
- MOLY, P. K., COOLEY, J. R., ZELTZER, S. L., YATSKIEVYCH, T. A. & ANTIN, P. B. 2016. Gastrulation EMT Is Independent of P-Cadherin Downregulation. *PloS one*, 11, e0153591.
- MOREL, A. P., LIEVRE, M., THOMAS, C., HINKAL, G., ANSIEAU, S. & PUISIEUX, A. 2008. Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One*, 3, e2888.
- MURRAY, S. A. & GRIDLEY, T. 2006. Snail family genes are required for left-right asymmetry determination, but not neural crest formation, in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 10300-10304.
- NIETO, M. A. 2001. The early steps of neural crest development. *Mechanisms of development*, 105, 27-35.
- NIETO, M. A. 2002. The snail superfamily of zinc-finger transcription factors. *Nature reviews. Molecular cell biology*, 3, 155-66.
- NIETO, M. A. 2013. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science*, 342, 1234850.
- NIETO, M. A. 2017. Context-specific roles of EMT programmes in cancer cell dissemination. *Nature cell biology*, 19, 416-418.
- NIETO, M. A., BENNETT, M. F., SARGENT, M. G. & WILKINSON, D. G. 1992. Cloning and developmental expression of Sna, a murine homologue of the Drosophila snail gene. *Development*, 116, 227-37.

REFERENCES

- NIETO, M. A. & CANO, A. 2012a. The epithelial-mesenchymal transition under control: global programs to regulate epithelial plasticity. *Semin Cancer Biol*, 22, 361-8.
- NIETO, M. A. & CANO, A. The epithelial-mesenchymal transition under control: global programs to regulate epithelial plasticity. *Seminars in cancer biology*, 2012b. Elsevier, 361-368.
- NIETO, M. A., HUANG, R. Y., JACKSON, R. A. & THIERY, J. P. 2016. EMT: 2016. *Cell*, 166, 21-45.
- NIETO, M. A., PATEL, K. & WILKINSON, D. G. 1996. In situ hybridization analysis of chick embryos in whole mount and tissue sections. *Methods Cell Biol*, 51, 219-35.
- NIETO, M. A., SARGENT, M. G., WILKINSON, D. G. & COOKE, J. 1994. Control of cell behavior during vertebrate development by Slug, a zinc finger gene. *Science*, 264, 835-9.
- NORRIS, R. A. & KERN, M. J. 2001. The identification of Prx1 transcription regulatory domains provides a mechanism for unequal compensation by the Prx1 and Prx2 loci. *J Biol Chem*, 276, 26829-37.
- OCANA, O. H., CORCOLES, R., FABRA, A., MORENO-BUENO, G., ACLOQUE, H., VEGA, S., BARRALLO-GIMENO, A., CANO, A. & NIETO, M. A. 2012. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prx1. *Cancer cell*, 22, 709-24.
- OCANA, O. H., COSKUN, H., MINGUILLON, C., MURAWALA, P., TANAKA, E. M., GALCERAN, J., MUNOZ-CHAPULI, R. & NIETO, M. A. 2017. A right-handed signalling pathway drives heart looping in vertebrates. *Nature*, 549, 86-90.
- PASTUSHENKO, I., BRISEBARRE, A., SIFRIM, A., FIORAMONTI, M., REVENCO, T., BOUMAHDI, S., VAN KEYMEULEN, A., BROWN, D., MOERS, V., LEMAIRE, S., DE CLERCQ, S., MINGUIJON, E., BALSAT, C., SOKOLOW, Y., DUBOIS, C., DE COCK, F., SCOZZARO, S., SOPENA, F., LANAS, A., D'HAENE, N., SALMON, I., MARINE, J. C., VOET, T., SOTIROPOULOU, P. A. & BLANPAIN, C. 2018. Identification of the tumour transition states occurring during EMT. *Nature*, 556, 463-468.
- PATEL, K., ISAAC, A. & COOKE, J. 1999. Nodal signalling and the roles of the transcription factors SnR and Pitx2 in vertebrate left-right asymmetry. *Current biology : CB*, 9, 609-12.
- PEINADO, H., OLMEDA, D. & CANO, A. 2007. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer*, 7, 415-28.
- PEINADO, H., QUINTANILLA, M. & CANO, A. 2003. Transforming growth factor beta-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions. *J Biol Chem*, 278, 21113-23.
- PEIRO, S., ESCRIVA, M., PUIG, I., BARBERA, M. J., DAVE, N., HERRANZ, N., LARRIBA, M. J., TAKKUNEN, M., FRANCI, C., MUNOZ, A., VIRTANEN, I., BAULIDA, J. & GARCIA DE HERREROS, A. 2006. Snail1 transcriptional repressor binds to its own promoter and controls its expression. *Nucleic Acids Res*, 34, 2077-84.
- PENA, J. T., SOHN-LEE, C., ROUHANIFARD, S. H., LUDWIG, J., HAFNER, M., MIHAILOVIC, A., LIM, C., HOLOCH, D., BERNINGER, P., ZAVOLAN, M. & TUSCHL,

- T. 2009. miRNA in situ hybridization in formaldehyde and EDC-fixed tissues. *Nature methods*, 6, 139-41.
- PERNAUTE, B., SPRUCE, T., RODRIGUEZ, T. A. & MANZANARES, M. 2011. MiRNA-mediated regulation of cell signaling and homeostasis in the early mouse embryo. *Cell Cycle*, 10, 584-91.
- PIJNENBORG, R., DIXON, G., ROBERTSON, W. B. & BROSENS, I. 1980. Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy. *Placenta*, 1, 3-19.
- POLLHEIMER, J., FOCK, V. & KNOFLER, M. 2014. Review: the ADAM metalloproteinases - novel regulators of trophoblast invasion? *Placenta*, 35 Suppl, S57-63.
- PURAM, S. V., TIROSH, I., PARIKH, A. S., PATEL, A. P., YIZHAK, K., GILLESPIE, S., RODMAN, C., LUO, C. L., MROZ, E. A., EMERICK, K. S., DESCHLER, D. G., VARVARES, M. A., MYLVAGANAM, R., ROZENBLATT-ROSEN, O., ROCCO, J. W., FAQUIN, W. C., LIN, D. T., REGEV, A. & BERNSTEIN, B. E. 2017. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, 171, 1611-1624 e24.
- REMBOLD, M., CIGLAR, L., YANEZ-CUNA, J. O., ZINZEN, R. P., GIRARDOT, C., JAIN, A., WELTE, M. A., STARK, A., LEPTIN, M. & FURLONG, E. E. 2014. A conserved role for Snail as a potentiator of active transcription. *Genes & development*, 28, 167-81.
- RIGGI, N., AGUET, M. & STAMENKOVIC, I. 2018. Cancer Metastasis: A Reappraisal of Its Underlying Mechanisms and Their Relevance to Treatment. *Annual review of pathology*, 13, 117-140.
- RODRIGUEZ-AZNAR, E., BARRALLO-GIMENO, A. & NIETO, M. A. 2013. Scratch2 prevents cell cycle re-entry by repressing miR-25 in postmitotic primary neurons. *J Neurosci*, 33, 5095-105.
- ROWE, R. G., LI, X. Y., HU, Y., SAUNDERS, T. L., VIRTANEN, I., GARCIA DE HERREROS, A., BECKER, K. F., INGVARSEN, S., ENGELHOLM, L. H., BOMMER, G. T., FEARON, E. R. & WEISS, S. J. 2009. Mesenchymal cells reactivate Snail1 expression to drive three-dimensional invasion programs. *The Journal of cell biology*, 184, 399-408.
- SALNIKOV, A. V., LIU, L., PLATEN, M., GLADKICH, J., SALNIKOVA, O., RYSCHICH, E., MATTERN, J., MOLDENHAUER, G., WERNER, J., SCHEMMER, P., BUCHLER, M. W. & HERR, I. 2012. Hypoxia induces EMT in low and highly aggressive pancreatic tumor cells but only cells with cancer stem cell characteristics acquire pronounced migratory potential. *PloS one*, 7, e46391.
- SARRIO, D., RODRIGUEZ-PINILLA, S. M., HARDISSON, D., CANO, A., MORENO-BUENO, G. & PALACIOS, J. 2008. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res*, 68, 989-97.
- SAVAGNER, P., KUSEWITT, D. F., CARVER, E. A., MAGNINO, F., CHOI, C., GRIDLEY, T. & HUDSON, L. G. 2005. Developmental transcription factor slug is required for effective re-epithelialization by adult keratinocytes. *Journal of cellular physiology*, 202, 858-66.

- SEFTON, M., SANCHEZ, S. & NIETO, M. A. 1998. Conserved and divergent roles for members of the Snail family of transcription factors in the chick and mouse embryo. *Development*, 125, 3111-21.
- SHAW, T. J. & MARTIN, P. 2016. Wound repair: a showcase for cell plasticity and migration. *Current opinion in cell biology*, 42, 29-37.
- SHIMONO, Y., ZABALA, M., CHO, R. W., LOBO, N., DALERBA, P., QIAN, D., DIEHN, M., LIU, H., PANULA, S. P., CHIAO, E., DIRBAS, F. M., SOMLO, G., PERA, R. A., LAO, K. & CLARKE, M. F. 2009. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell*, 138, 592-603.
- SMYTH, G. K. 2005. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer.
- SUGIYAMA, M., HASEGAWA, H., ITO, S., SUGIYAMA, K., MAEDA, M., AOKI, K., WAKABAYASHI, T., HAMAGUCHI, M., NATSUME, A. & SENGU, T. 2015. Paired related homeobox 1 is associated with the invasive properties of glioblastoma cells. *Oncology reports*, 33, 1123-30.
- TAUBE, J. H., HERSCHKOWITZ, J. I., KOMUROV, K., ZHOU, A. Y., GUPTA, S., YANG, J., HARTWELL, K., ONDER, T. T., GUPTA, P. B., EVANS, K. W., HOLLIER, B. G., RAM, P. T., LANDER, E. S., ROSEN, J. M., WEINBERG, R. A. & MANI, S. A. 2010. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 15449-54.
- THEVENEAU, E. & MAYOR, R. 2012. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Developmental biology*, 366, 34-54.
- THIERY, J. P., ACLOQUE, H., HUANG, R. Y. & NIETO, M. A. 2009. Epithelial-mesenchymal transitions in development and disease. *Cell*, 139, 871-90.
- TIROSH, I., IZAR, B., PRAKADAN, S. M., WADSWORTH, M. H., 2ND, TREACY, D., TROMBETTA, J. J., ROTEM, A., RODMAN, C., LIAN, C., MURPHY, G., FALLAHI-SICHANI, M., DUTTON-REGESE, K., LIN, J. R., COHEN, O., SHAH, P., LU, D., GENSHAFT, A. S., HUGHES, T. K., ZIEGLER, C. G., KAZER, S. W., GAILLARD, A., KOLB, K. E., VILLANI, A. C., JOHANNESSEN, C. M., ANDREEV, A. Y., VAN ALLEN, E. M., BERTAGNOLLI, M., SORGER, P. K., SULLIVAN, R. J., FLAHERTY, K. T., FREDERICK, D. T., JANE-VALBUENA, J., YOON, C. H., ROZENBLATT-ROSEN, O., SHALEK, A. K., REGEV, A. & GARRAWAY, L. A. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, 352, 189-96.
- TRAN, D. D., CORSA, C. A., BISWAS, H., AFT, R. L. & LONGMORE, G. D. 2011. Temporal and spatial cooperation of Snail1 and Twist1 during epithelial-mesenchymal transition predicts for human breast cancer recurrence. *Mol Cancer Res*, 9, 1644-57.
- TRAN, H. D., LUITEL, K., KIM, M., ZHANG, K., LONGMORE, G. D. & TRAN, D. D. 2014. Transient SNAIL1 expression is necessary for metastatic competence in breast cancer. *Cancer research*, 74, 6330-40.
- TSAI, J. H., DONAHER, J. L., MURPHY, D. A., CHAU, S. & YANG, J. 2012. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer cell*, 22, 725-36.

- VALDES, F., ALVAREZ, A. M., LOCASCIO, A., VEGA, S., HERRERA, B., FERNANDEZ, M., BENITO, M., NIETO, M. A. & FABREGAT, I. 2002. The epithelial mesenchymal transition confers resistance to the apoptotic effects of transforming growth factor Beta in fetal rat hepatocytes. *Mol Cancer Res*, 1, 68-78.
- VEGA, S., MORALES, A. V., OCANA, O. H., VALDES, F., FABREGAT, I. & NIETO, M. A. 2004. Snail blocks the cell cycle and confers resistance to cell death. *Genes & development*, 18, 1131-43.
- VILLAREJO, A., CORTES-CABRERA, A., MOLINA-ORTIZ, P., PORTILLO, F. & CANO, A. 2014. Differential role of Snail1 and Snail2 zinc fingers in E-cadherin repression and epithelial to mesenchymal transition. *J Biol Chem*, 289, 930-41.
- WAGNER, D. E., WEINREB, C., COLLINS, Z. M., BRIGGS, J. A., MEGASON, S. G. & KLEIN, A. M. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*.
- WU, Y. & ZHOU, B. P. 2010. Snail: More than EMT. *Cell adhesion & migration*, 4, 199-203.
- XIE, T., WANG, Y., DENG, N., HUANG, G., TAGHAVIFAR, F., GENG, Y., LIU, N., KULUR, V., YAO, C., CHEN, P., LIU, Z., STRIPP, B., TANG, J., LIANG, J., NOBLE, P. W. & JIANG, D. 2018. Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. *Cell reports*, 22, 3625-3640.
- XIE, Z., HU, S., BLACKSHAW, S., ZHU, H. & QIAN, J. 2010. hPDI: a database of experimental human protein-DNA interactions. *Bioinformatics*, 26, 287-9.
- XU, X., TAN, X., TAMPE, B., SANCHEZ, E., ZEISBERG, M. & ZEISBERG, E. M. 2015. Snail Is a Direct Target of Hypoxia-inducible Factor 1alpha (HIF1alpha) in Hypoxia-induced Endothelial to Mesenchymal Transition of Human Coronary Endothelial Cells. *The Journal of biological chemistry*, 290, 16653-64.
- XU, Y., LEE, D. K., FENG, Z., BU, W., LI, Y., LIAO, L. & XU, J. 2017. Breast tumor cell-specific knockout of Twist1 inhibits cancer cell plasticity, dissemination, and lung metastasis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 11494-11499.
- YANG, F., HUANG, X. R., CHUNG, A. C., HOU, C. C., LAI, K. N. & LAN, H. Y. 2010. Essential role for Smad3 in angiotensin II-induced tubular epithelial-mesenchymal transition. *The Journal of pathology*, 221, 390-401.
- YANG, M. H., WU, M. Z., CHIOU, S. H., CHEN, P. M., CHANG, S. Y., LIU, C. J., TENG, S. C. & WU, K. J. 2008. Direct regulation of TWIST by HIF-1alpha promotes metastasis. *Nature cell biology*, 10, 295-305.
- YE, X., TAM, W. L., SHIBUE, T., KAYGUSUZ, Y., REINHARDT, F., NG EATON, E. & WEINBERG, R. A. 2015. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature*, 525, 256-60.
- YOKOYAMA, K., KAMATA, N., FUJIMOTO, R., TSUTSUMI, S., TOMONARI, M., TAKI, M., HOSOKAWA, H. & NAGAYAMA, M. 2003. Increased invasion and matrix metalloproteinase-2 expression by Snail-induced mesenchymal transition in squamous cell carcinomas. *International journal of oncology*, 22, 891-8.

REFERENCES

- YU, M., BARDIA, A., WITTNER, B. S., STOTT, S. L., SMAS, M. E., TING, D. T., ISAKOFF, S. J., CICILIANO, J. C., WELLS, M. N., SHAH, A. M., CONCANNON, K. F., DONALDSON, M. C., SEQUIST, L. V., BRACHTEL, E., SGROI, D., BASELGA, J., RAMASWAMY, S., TONER, M., HABER, D. A. & MAHESWARAN, S. 2013. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*, 339, 580-4.
- ZHANG, D., SHI, Z., LI, M. & MI, J. 2014. Hypoxia-induced miR-424 decreases tumor sensitivity to chemotherapy by inhibiting apoptosis. *Cell death & disease*, 5, e1301.
- ZHANG, J., CHENG, Q., ZHOU, Y., WANG, Y. & CHEN, X. 2013. Slug is a key mediator of hypoxia induced cadherin switch in HNSCC: correlations with poor prognosis. *Oral oncology*, 49, 1043-50.
- ZHANG, W., SHI, X., PENG, Y., WU, M., ZHANG, P., XIE, R., WU, Y., YAN, Q., LIU, S. & WANG, J. 2015. HIF-1alpha Promotes Epithelial-Mesenchymal Transition and Metastasis through Direct Regulation of ZEB1 in Colorectal Cancer. *PloS one*, 10, e0129603.
- ZHENG, X., CARSTENS, J. L., KIM, J., SCHEIBLE, M., KAYE, J., SUGIMOTO, H., WU, C. C., LEBLEU, V. S. & KALLURI, R. 2015. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature*, 527, 525-530.